

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 08:13:15 ; Search time 382.871 Seconds  
(without alignments)  
10794.564 Million cell updates/sec

Title: US-09-729-674-1\_COPY\_132\_1265

Perfect score: 1134  
Sequence: 1 gggcgcccaatacgaacg.....tggaaactggagagaccagc 1134

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1133.2	99.9	3871	9	US-09-729-674-1
2	850	75.0	1810	12	US-10-108-260A-2059
3	455.4	40.2	2347	10	US-09-562-832-113
4	455.4	40.2	2347	12	US-10-439-388-64
5	455.4	40.2	2443	13	US-10-094-749-724
6	196.6	17.3	398	12	US-10-242-535A-8221
7	154.8	13.7	2600	14	US-10-045-815-3
8	154.8	13.7	2829	14	US-10-045-815-1
9	153.8	13.6	892	13	US-10-027-632-161474
10	153.8	13.6	892	13	US-10-027-632-161475
11	153.8	13.6	892	13	US-10-027-632-161476
12	153.8	13.6	892	14	US-10-027-632-161475
13	153.8	13.6	892	14	US-10-027-632-161475
14	153.8	13.6	892	14	US-10-027-632-161476
15	147.4	13.0	2322	14	US-10-045-815-7

16	147.4	13.0	2416	14	US-10-045-815-5
17	145.4	12.8	564	13	US-10-027-632-114858
18	145.4	12.8	564	14	US-10-027-632-114858
19	128.4	11.3	3396	13	US-10-094-749-921
20	126	11.1	618	10	US-09-879-536-215
21	109.2	9.6	2685	14	US-10-071-766-86
22	107.2	9.5	286	12	US-10-242-535A-9365
23	103.4	9.1	150	12	US-10-242-535A-1775
24	97.8	8.6	2753	10	US-09-728-952-28
25	97.8	8.6	2753	10	US-09-728-952-51
26	87	7.7	349	9	US-09-925-303-395
27	73	6.4	247	12	US-10-242-535A-4919
28	69.6	6.1	181	13	US-10-023-386-16058
29	69.6	6.1	549	13	US-10-023-386-2358
30	62.4	5.5	3897	13	US-10-412-000-7
31	62.4	5.5	3897	14	US-10-021-657-7
32	55.8	4.9	2404	12	US-10-260-238-26
33	53.8	4.7	2305	12	US-10-094-886-23
34	53.4	4.7	1104	15	US-10-156-761-6371
35	53.4	4.7	9025608	15	US-10-156-761-1
36	53.2	4.7	1374	15	US-10-158-761-7192
37	52.6	4.6	1906	13	US-10-412-000-1
38	52.6	4.6	1906	14	US-10-021-657-1
39	52.2	4.6	986	15	US-10-156-761-3002
40	52.2	4.6	1413	15	US-10-156-761-4744
41	52.2	4.6	9025608	15	US-10-156-761-1
42	52	4.6	3624	11	US-09-988-462-6
43	51.6	4.6	451	13	US-10-029-386-20388
44	51.6	4.6	2793	15	US-10-158-761-2591
45	51.4	4.5	3275	13	US-10-027-632-114407

#### ALIGNMENTS

RESULT 1

US-09-729-674-1  
; Sequence 1, Application US/09729674  
; Patent No. US20010039335A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: LaVallie, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steininger II, Robert J.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fectel, Kim  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: 6055-64X  
; CURRENT APPLICATION NUMBER: US/09/729,674  
; CURRENT FILING DATE: 2000-12-04  
; PRIOR APPLICATION NUMBER: 09/539,330  
; NUMBER OF SEQ ID NOS: 283  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 3871  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-729-674-1

Query Match 99.9%; Score 1133.2; DB 9; Length 3871;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCGGCCCAATACGAACGTACAGCTTCGCGAGCTTCCCGAGCGAGCGAGTGCGG 60





301	ACGAGTGGCCCTCGAGCTGCGGCTCTTCGGCGGCTCTGAGAGGAGCGCGCTGCTGC	360
290	AGCGCTGCAAGAGGGCGCTGCGAGCTTCCGCACTCCGAGGCCAGCGCGAGGTGCTGG	349
361	GGCGCTGCAAGCGGAGCGCTTCGCGCTTCCAGTGCCTTACCGCGCGGAGCTGCTGC	420
350	CGGACTTCCAGCGCGCGAGCCCTACAAAGTTCCTGAGTTCCTGCTTCAAGGCAATA	409
421	GTGACTTCCAGAGCGCGCTGCGCTACCACTACCTGACTACCGCTGTTCAGAGGTAACC	480
410	ATCTCCCAAGGCCATCGCGCTGCTCACACCTTCTACTGAAGCATCCTGATGACGAAA	469
481	GGCTGAGAAGCGGTGGCGGCGGCTACACCTTCTCCAGAGGAACCGAAGCAGGAGC	540
470	TGATGAAGAGGAACATGGCATATTATAA---GAGCGTGCCTGGTGCAGGACTCAATTA	526
541	TGACCGCCAAGTATCTCAACTACTATCAGGGATGCTGACGTGCGCAGGAGTCCCTCA	600
527	AAGACCTGGAACCAAGTCAATATGAAGCCTGTTATCCGACAGTGGGGCATCAACG	586
601	CGGACCTAGAGGCCACGCGCTACGAGCGCGTGTCTCGGGCTGTGAAGCTCTTACAACA	660
587	GTGAGAACTGGAGAACATCCATCACAGACATGAGGTGGCCCTTCCCGACTTCTTCAAAG	646
661	GCGGGATTTCGCGAGCAGCAGGAGGACATGGAGGGGCCCTTGTCCAGAGTACCTGGCAG	720
647	CCTTTACGAGTGTCTGCGAGCGCTGCGAGGGTTCAGGGAGATCAAGGACTTCMAGGATT	706
721	TCTTGCGCGGTGCCTGGCGCGCTGTGAAGGGGCCCATGAGCAGGTGGACTTCAAGGACT	780
707	TCCTACCTTCCATAGCAGATCAATTATGTAGAAGTCTGGAATGCAAAATACAGTGTGAAG	766
781	TCTACCGGCCATAGCAGATCTCTTTCGAGAGTCCCTGCAGTCAAGGTGGACTGTGAGG	840
767	AGAACCTCACCCAGTTATAGGAGGCTATCCGGTTGAGAAAATTTGTGGCTACCATGTATC	826
841	CCAAATTGACCCCAATGTGGTGGGTACTTCTGTGGACAAGTTCTGTGGCCACCATGTACC	900
827	ATTACTTTCGAGTTGCGCTATTATTAAGTTGAACACCTGAAGATGCAGCCCCCTGTGCAG	886
901	ACTACCTGCGAGTTGGCTACTATTAAGTTGAATGATGTGGCGAGGCTGCCCGCAGCGCG	960
887	TCAGCTATCTGCTCTTTGATCAGAATGACAAGGTCATGACGACAGAACCTGGTGTAATTACC	946
961	CCAGCTACATGCTCTTTCGACCCCAAGACAGCGTCATGACGACAGAACCTGGTGTAATTACC	1020
947	AGTACCAAGGACACTTGGGGCTCTCGATGACGACTTCCAGCCCGACACCTGGAAGCAG	1006
1021	GGTTTCCACGGGCTCGCTGGGGCGCTGGAAGAGGAGACTTTCAGCCCCGGGAGGAGGCCA	1080
1007	TTCAAGTTCTTAAATGTGACCACACTCCAGAAAGGAGCTGTATGACTTTGTCTAAGGAAAATA	1066
1081	TGCTCTACCAACACCGACGCCGAGCTGCGGAGCTGCTGGAGTTTCAACCATGTATACC	1140
1067	TAAATGGATGATGAGGGAGAAAGTTGTGGA	1097
1141	TGCAATCAGATGATGAGATGAGGCTGGAGGA	1171

```

US-10-439-388-64
? Sequence 64, Application US/10439388
? Publication No. US20030228617A1
? GENERAL INFORMATION:
? APPLICANT: Aure, Thomas M
? APPLICANT: Olsen, Nancy J
? TITLE OF INVENTION: Method for Predicting Autoimmune Disease
? FILE REFERENCE: 1242/68
? CURRENT APPLICATION NUMBER: US/10/439,388
? CURRENT FILING DATE: 2003-05-16
? PRIOR APPLICATION NUMBER: US 60/381,055
? PRIOR FILING DATE: 2002-05-16

```

Query Match	40.2%;	Score 455.4;	DB 10;	Length 2347;
Best Local Similarity	65.1%;	Pred.No. 2.7e-125;		
Matches 723; Conservative	12;	Mismatches 368;	Indels 18;	Gaps 3
<hr/>				
Qy	5	GGGCCCAATACGAACGCCTACAGCTTCGCGAGCTTCCCACGGACAGAGTGATGCCGCTCG	64	
Dd	61	GGGCGCAGTACAGAAGTACAGCTTCGGGGCTTCCGCGCGAGGACCTGATGCGCTGG	120	
<hr/>				
Qy	65	AGTCGGCTTACCGGACCGCGCTGGACAAGTACAGCGGCGAGCACCTTGGGCCGAGAGCSIKG	124	
<hr/>				
Dd	121	CGCGCGCTACCGGACCGCTCTGGAGCAGTA CGAGGAGAGAGCTGGCCGAGAGCGCGC	180	
<hr/>				
Qy	125	GCTACTGGAGATCAGCTTCGCGCTGCACCGTTGCTGCGACACGAGGCGCTTCTGCC	184	
Dd	181	GCTACTGGAGCGCGGCTCGGCTGCACCGGCTCTGTGCGCACGAGGAGGCTTCTGCC	240	
<hr/>				
Qy	185	ACGGCACTGCAGCG-----CGCGCGCAGCGCCGAGCCCGCGCGCTCGCCA	235	
<hr/>				
Dd	241	ACGCCAATGCAGCGGCCCGCGCCCGCGGCCAAGCCCCATCCGACGCGCGCGCGAG	300	
<hr/>				
Qy	236	GGTA-----TCCGAGCTCGCGCTCTTCGGGGGCGCTGTGCGCCGCGGGACTGCCTCA	289	

```

; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 64
; LENGTH: 2347
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-439-388-64

Query Match      40.2%; Score 455.4; DB 12; Length 2347;
Best Local Similarity 65.1%; Pred. No. 2.7e-125;
Matches 723; Conservative 2; Mismatches 368; Indels 18; Gaps 3;

QY 5 GCGCCCAATACGACGCTCAGCTTCGACGCTTCCACGGGACGAGCTGCGCTCG 64
DB 61 GGGCGAGTACGAGAGTACAGCTTCGCGGCTTCCGCGGAGGACCTGATGCGCTGG 120
QY 65 AGTGGGCTACCGGACGCGCTCGACAAATGATACAGCGGCGAGCACTGGGCGGAGAGCTG 124
DB 121 CCGGGGCTACGGGACGCTCTGAGCAGTACGAGGAGAGAGCTGGCGGAGAGCGGC 180
QY 125 GCTACCTGGAGATCAGCTCGGCTGACCGCTTGTGCGCGACAGCGAGGCTTCTGCG 184
DB 181 GCTACCTGGAGGCGGCGCTGCGGCTGCACCGCTTCTGCGCGACAGCGAGGCTTCTGCG 240
QY 185 ACCGCAACTGCAGCG-----CCGCGCGGACGCGGCGGCGCGGCGCTCGCGA 235
DB 241 ACGCAACTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
QY 236 GCTA-----TCCGAGCTGGCTCTTCGGGGGCTGCTGCGCGCGCGCGCGCTG 289
DB 301 ACGAGTGGGCTGCGAGCTGCGGCTCTTCGCGCGCTTCCAGTGCGCTACCGCGCGCGCTG 360
QY 290 ACGCTGCAAGCAGGCGCTGCGAGCTTCCGCGCGCTTCCAGTGCGCTACCGCGCGCGCTG 349
DB 361 GCGCTGCAAGCAGGCGCTGCGGCTTCCAGTGCGCTACCGCGCGCGCGCTG 420
QY 350 CGGACTTCCAGCGCGCGGCGCGCTTCCAGTGCGCTTCCAGTGCGCTTCCAGTGCGCT 409
DB 421 GTGACTTCCAGCGCGCGCTGCGGCTTCCAGTGCGCTTCCAGTGCGCTTCCAGTGCGCT 480
QY 410 ATCTCCCAAGCATGCGCGCTGCTCAGACTTCTGCTGAGGCTTCTGAGGCTTCTGAGGAA 469
DB 481 GCGTGGAGAGCGGCTGCGGCGCGCTTCCAGTGCGCTTCCAGTGCGCTTCCAGTGCGCT 540
QY 470 TGATGAAGAGAAATGCGCATATTATAA---GAGCTGCTGCTGCGGAGCTTACATTA 526
DB 541 TGACCGCAAGTATCTCAACTACTATCAGGCGGATGCTGCGGCTGCGGCGGAGCTTCCGCTCA 600
QY 527 AAGACCTGGAACCAAGCTCATGAAAGCTGTTTCATCCGAGCAGTGGGCGCATACAAAG 586
DB 601 CGGACCTAGAGCGCGCGCGCTTACAGGCGCTGTTCTCCGCGCTGTAAGCTTACAA 660
QY 587 GTGAGAACTGGAGAACATCCATCAGACATGAGAGTGGGCTTCCGAGCTTCTTCAAG 646
DB 661 GCGGGAATTCGCGAGCAGCAGGAGGACATGAGGCGGCGCTTCTGAGTACCTGCGAG 720
QY 647 CTTTACAGTGTCTGCGAGCTTCCAGGCTTCCAGGAGATCAAGAGCTTCAAGATT 706
DB 721 TCTTTCGCGGCTGCGCGGCTGTAAGGCGGCGCTGTAAGGCGGCGCTTCAAGAGCT 780
QY 707 TCTACTTTCATAGCAGATCATTTATGTAAGTTCTGGAATGCAAAATACAGTGTGAAG 766
DB 781 TCTACCGCGCATAGCAGATCTTTTCAGAGTCCCTGAGTCAAGGTGAGCTGTGAGG 840
QY 767 AGAAGCTCAGCGGATTAAGAGCTATCCGTTGAGAAATTTTGGCTACCATGATTC 826
DB 841 CCAATTTGACCCCAATGCGGCTGTAAGGCGGCGCTTCTGCGGAGCTTCTGCGGCGGCTG 900
QY 827 ATTACTTCAGTTTTCGCTTATTAAGTTTGAACGCTTGAAGATGCAAGCTTCCGCTGTCAG 886
DB 901 ACTACTTCAGTTTTCGCTTATTAAGTTTGAAGTATGATGTCGCGGAGGCTTCCGCGGCGG 960
QY 887 TCAGTATCTGCTCTTTGATCAGAAATGACAAGGTGATGCGAGCAACCTGCTGTATTACC 946

RESULT 5
US-10-094-749-724
; Sequence 724, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 724
; LENGTH: 2443
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-724

Query Match      40.2%; Score 455.4; DB 13; Length 2443;
Best Local Similarity 65.1%; Pred. No. 2.8e-125;
Matches 723; Conservative 2; Mismatches 368; Indels 18; Gaps 3;

QY 5 GCGCCCAATACGACGCTCAGCTTCCGAGCTTCCACGGGACGAGCTGATGCGCTCG 64
DB 517 GGGCGGAGTACGAGAGTACAGCTTCCGCGGCTTCCGCGGAGGACCTGATGCGCTGG 576
QY 65 AGTGGGCTTACCGGACGCGCTGGAAGATGACAGCGGCGAGCACTGGGCGGAGAGCTG 124
DB 577 CCGGCGGCTACGGGACGCTCTGGAGCAGTACGAGGAGAGAGCTGGGCGGAGAGCGGC 636
QY 125 GCTACTGAGATACGCTTGGCTGCGGCTTCTGCGGAGAGAGGCTTCTGCG 184
DB 637 GCTACTGAGGCGGCTGCGGCTTCTGCGGAGAGAGGCTTCTGCG 696
QY 185 ACGCAACTGCGAGCG-----CCGCGCGCGAGCGCGCGCGCGCTTCCGCA 235
```





Db 1047 CAATGACGAGTGTGATGAACCAAAATTTGGCTATTA 1082

## RESULT 9

US-10-027-632-161474/c  
; Sequence 161474, Application US/10027632  
; Publication No. US20030204075A9

## GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 161474

; LENGTH: 892

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-161474

## Query Match

Best Local Similarity 13.6%; Score 153.8; DB 13; Length 892;

Best Local Similarity 97.5%; Pred. No. 3e-35; 3; Indels 0; Gaps 0;

Matches 155; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 394 TACTTCAAGGCAATAATCTCCCAAGCCATCGCGCTGCTCAGACCTTTCTACTGAAG 453

Db 228 TCCCTTTAGGCAATAATCTCCCAAGCCATCGCGCTGCTCAGACCTTTCTACTGAAG 169

QY 454 CATCTGATGACGAATGATGAGGAAACATGCGCATATTATTAAGAGCCTGCTGTGCC 513

Db 168 CATCTGATGAGAAATGATGAGGAAACATGCGCATATTATTAAGAGCCTGCTGTGCC 109

QY 514 GAGGACTACATTAAAGACCTGGAACCAAGTCATATGAA 552

Db 108 GAGGACTACATTAAAGACCTGGAACCAAGTCATATGAA 70

## RESULT 10

US-10-027-632-161475/c

; Sequence 161475, Application US/10027632

; Publication No. US20030204075A9

## GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 161475

; LENGTH: 892

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-161475

## Query Match

Best Local Similarity 13.6%; Score 153.8; DB 13; Length 892;

Best Local Similarity 97.5%; Pred. No. 3e-35; 3; Indels 0; Gaps 0;

Matches 155; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 394 TACTTCAAGGCAATAATCTCCCAAGCCATCGCGCTGCTCAGACCTTTCTACTGAAG 453

Db 228 TCCCTTTAGGCAATAATCTCCCAAGCCATCGCGCTGCTCAGACCTTTCTACTGAAG 169

QY 454 CATCTGATGACGAATGATGAGGAAACATGCGCATATTATTAAGAGCCTGCTGTGCC 513

Db 168 CATCTGATGAGAAATGATGAGGAAACATGCGCATATTATTAAGAGCCTGCTGTGCC 109

QY 514 GAGGACTACATTAAAGACCTGGAACCAAGTCATATGAA 552

Db 108 GAGGACTACATTAAAGACCTGGAACCAAGTCATATGAA 70

## RESULT 11

US-10-027-632-161476/c

; Sequence 161476, Application US/10027632

; Publication No. US20030204075A9

## GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 161476

; LENGTH: 892

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-161476

## Query Match

Best Local Similarity 13.6%; Score 153.8; DB 13; Length 892;

Best Local Similarity 97.5%; Pred. No. 3e-35; 3; Indels 0; Gaps 0;

Matches 155; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 394 TACTTCAAGGCAATAATCTCCCAAGCCATCGCGCTGCTCAGACCTTTCTACTGAAG 453

Db 228 TCCCTTTAGGCAATAATCTCCCAAGCCATCGCGCTGCTCAGACCTTTCTACTGAAG 169

QY 454 CATCTGATGACGAATGATGAGGAAACATGCGCATATTATTAAGAGCCTGCTGTGCC 513

Db 168 CATCTGATGAYCAATGATGAAGAGGAACATGGCATATATATAAGAGCCTGCTGTGTC 109  
Qy 514 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 552  
Db 108 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 70

RESULT 12  
US-10-027-632-161474/c  
; Sequence 161474, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 161474  
; LENGTH: 892  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-161474

Query Match 13.6%; Score 153.8; DB 14; Length 892;  
Best Local Similarity 97.5%; Pred. No. 3e-35;  
Matches 155; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
Qy 394 TACTTCAAGGCAATAATCTCCCAAGCCATCCCGCTGCTCAGCTTCTACTGAAG 453  
Db 228 TCCCTTTAGGCAATAATCTCCCAAGCCATCCCGCTGCTCAGCTTCTACTGAAG 169  
Qy 454 CATCTGATGACGAAATGATGAAGAGGAACATGGCATATATATAAGAGCCTGCTGTGTC 513  
Db 168 CATCTGATGAYCAATGATGAAGAGGAACATGGCATATATATAAGAGCCTGCTGTGTC 109  
Qy 514 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 552  
Db 108 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 70

RESULT 13  
US-10-027-632-161475/c  
; Sequence 161475, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 161475  
; LENGTH: 892  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-161475  
Query Match 13.6%; Score 153.8; DB 14; Length 892;  
Best Local Similarity 97.5%; Pred. No. 3e-35;  
Matches 155; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
Qy 394 TACTTCAAGGCAATAATCTCCCAAGCCATCCCGCTGCTCAGCTTCTACTGAAG 453  
Db 228 TCCCTTTAGGCAATAATCTCCCAAGCCATCCCGCTGCTCAGCTTCTACTGAAG 169  
Qy 454 CATCTGATGACGAAATGATGAAGAGGAACATGGCATATATATAAGAGCCTGCTGTGTC 513  
Db 168 CATCTGATGAYCAATGATGAAGAGGAACATGGCATATATATAAGAGCCTGCTGTGTC 109  
Qy 514 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 552  
Db 108 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 70

RESULT 14  
US-10-027-632-161476/c  
; Sequence 161476, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 161476  
; LENGTH: 892  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-161476

Query Match 13.6%; Score 153.8; DB 14; Length 892;  
Best Local Similarity 97.5%; Pred. No. 3e-35;  
Matches 155; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
Qy 394 TACTTCAAGGCAATAATCTCCCAAGCCATCCCGCTGCTCAGCTTCTACTGAAG 453  
Db 228 TCCCTTTAGGCAATAATCTCCCAAGCCATCCCGCTGCTCAGCTTCTACTGAAG 169  
Qy 454 CATCTGATGACGAAATGATGAAGAGGAACATGGCATATATATAAGAGCCTGCTGTGTC 513

```

Db 168 CATCTGATGAYGAAATGATGAAGAGAACATGGCATATTATAAGAGCCTGGCTGGTGC 109
Qy 514 GAGGACTACATTAAGACCTGGAAACCAAGTCATATGAA 552
Db 108 GAGGACTACATTAAGACCTGGAAACCAAGTCATATGAA 70

RESULT 15
US-10-045-815-7
; Sequence 7, Application US/10045815
; Publication No. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwah, Renu
; APPLICANT: Sugihara, Takashi
; APPLICANT: Ohide, Akiko
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045,815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JF00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118806
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2322
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(1637)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2282
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-7

Query Match 13.0%; Score 147.4; DB 14; Length 2322;
Best Local Similarity 54.6%; Pred. No. 4,1e-33;
Matches 421; Conservative 0; Mismatches 316; Indels 34; Gaps 5;

Qy 200 CGCGCGCGCAGCCGAGCC-CGCGCGCGCGCTCGCAGCTATCCGAGCTGCGCCTCTTC 258
Db 288 CCGAGACCCAGCTGAGCCAGACCCGCGCGCGCGCTCGCAGCTATCCGAGCTGCGCCTCTTC 347
Qy 259 GGGGGCTGTGCGCGCGCGCACTGCTCAAGCGCTGCAAGCAGGCGCTGCCAGCCTTC 318
Db 348 GGAGCGGTGTGCGCGCGCTGCTCAAGCGCTGCTCGCGCGCTGCTCGCGCGCTGCTCGCCAC 407
Qy 319 CGCCAGTCCAGCCAGCCGCGAGGTGTGCGGAGCTTCCAGCGCGCGAGCCCTACAG 378
Db 408 TTGCTG-----AGTGAAGAACTGGACCTGGAGTCAACAGCGGAGCCCGGTACAC 458
Qy 379 TTCTGTGAGTTCGGTTACTTCAAGGCAATAATCTCCCCAAAGCCATCGCGCTGTCTAC 438
Db 459 TACCTGAGGTGCGCTATTTCAGATAAACAAGCTGGAGAAAGCTGTGGTGGGACAC 518
Qy 439 ACCTTTCTACTGAAGCATCTGTATGACGAATGATGAAGAGGAACATGGCATATTATAAG 498
Db 519 ACCTTCTTTTGGGCAATCTGTAGCACATGGAGATGGGAGAACCTCGACTATTACCA 578
Qy 499 AGCTGTGCTGGTG--CCGAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAAAGC 555
Db 579 ACCATGTCTGGGTGAAGAGGAGGACATTCAGGGATCTCGAGGCCAAGCCCATATGCAT 638
Qy 556 CTGTTCATCCAGCAGTGGGGCATACACCGGTGAGAACTGGAGAACATCCATCACAGAC 615
Db 639 GAGTTTCGGCTGGGGTACGACTCTACTCAGAGGAGAAAGCCACAGGAAGCTGTGCCCCAC 698
Qy 616 ATGAGCTGGCCCTTCCGACTTCTTCAAGCCCTTTTACAGTGTCTCGCAGCCTGGAG 675

```

```

Db 699 CTGGAGGGCGCACTGCAAGAGTACTTTTGTGGCCGATGAGAGTGGCGCTCTGCGAA 758
Qy 676 GGTTCAGGGAGATCAAGGACTTCAA-----GGATTTCTACCTTTCC 717
Db 759 GGGCCCTATGACTACGACGGCTACAACTACCTAGACTACAGCGGTGACTCTTCCAGGCC 818
Qy 718 ATAGCAGATCATTTATGTAGAAGTTCTGGAATGCAAAATACAGTGTGAAGAAACCT---C 774
Db 819 ATCAGAGATCATTTACGTCAGGTCCCTCACTGTAAGCAGAACTGTGTCAAGAGCTGGCT 878
Qy 775 ACCCCAGTTATAGGAGGCTATCCGGTTGAGAAATTTGGGTACCATGTATCATTTACTTG 834
Db 879 TCCCAACCAAGTAGGAAAAGCCCTTTGAAGACTTTCCTCCCTTACACTATAATTACCTA 938
Qy 835 CAGTTTGCTCTATTATAAGTGTGAACGACCTGAAGAAATGACGCCCTGTGCGAGTCAGCTAT 894
Db 939 CAGTTTGCTCTACTACACATTTGGGAACCTATACACAAGCTATTGAATGTGCCAAGACCTAC 998
Qy 895 CTGCTCTTTGATCAGAAATGACAAGTCAATGACGAGAACCTGGTGGTATTATAC 945
Db 999 CTCCTCTTCTTTCCCAATGATGAGGTGATGCCACAGAACTCTGGCTTATTAC 1049

```

Search completed: January 30, 2004, 19:07:44  
Job time : 395.871 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 08:05:34 ; Search time 63.9336 Seconds  
(without alignments)  
7828.876 Million cell updates/sec

Title: US-09-729-674-1\_COPY\_132\_1265

Perfect score: 1134

Sequence: 1 gggcgcccaatacgaacg.....tgggaactggaggagaccagc 1134

Scoring table: IDENTITY NUC

Gapop 10\_0, Gapext 1.0

Searched: 569978 seqs, 220591566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	126	11.1	618	3	US-09-328-111-215
C 2	54.4	4.8	30001	1	US-08-125-468-1
C 3	54.4	4.8	30001	2	US-08-474-933-1
4	52	4.6	3624	1	US-07-951-715A-6
5	52	4.6	3624	2	US-08-459-448A-6
6	52	4.6	3624	3	US-08-459-595A-6
7	52	4.6	3624	3	US-08-459-504B-6
8	52	4.6	3624	3	US-08-459-444-6
9	52	4.6	3624	3	US-09-053-549-7
10	52	4.6	3624	4	US-09-547-422-6
11	50.6	4.5	942	4	US-09-252-991A-3976
12	50.6	4.5	1521	4	US-09-252-991A-4046
13	49.8	4.4	13942	3	US-09-105-537-30
14	49.8	4.4	36778	3	US-09-105-537-5
15	49.8	4.4	38506	3	US-09-320-878-19
16	49.8	4.4	38506	4	US-09-141-908-1
17	49.8	4.4	38506	4	US-09-657-440-19
18	49.2	4.3	2481	1	US-08-324-243-35
19	49.2	4.3	2481	2	US-08-532-390-35
20	49.2	4.3	2481	3	US-08-717-294-35
21	49.2	4.3	2481	5	PCR-US95-11511-35
22	49.2	4.3	2571	4	US-09-552-950-4
23	49	4.3	5228	4	US-09-428-711A-15
24	48.4	4.3	1162	2	US-08-726-306A-52
25	48.2	4.3	414	4	US-09-252-991A-2183
26	48.2	4.3	1635	4	US-09-252-991A-2493
27	48.2	4.3	1365	4	US-09-252-991A-9230

28	48.2	4.3	2805	4	US-09-252-991A-9064	Sequence 9064, Ap
C 29	48	4.2	567	4	US-09-252-991A-13282	Sequence 13282, A
30	48	4.2	1425	4	US-09-252-991A-12820	Sequence 12820, A
31	48	4.2	1428	4	US-09-252-991A-12677	Sequence 12677, A
C 32	48	4.2	1611	4	US-09-252-991A-13122	Sequence 13122, A
C 33	48	4.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 34	48	4.2	4411529	3	US-09-103-840A-1	Sequence 1, Appli
35	47.8	4.2	1155	4	US-09-252-991A-1572	Sequence 1572, Ap
36	47.8	4.2	4239	4	US-09-252-991A-1619	Sequence 1619, Ap
C 37	47.8	4.2	4353	4	US-09-252-991A-1517	Sequence 1517, Ap
38	47.5	4.2	1288	1	US-08-440-856A-9	Sequence 9, Appli
39	47.2	4.2	1637	4	US-09-252-991A-15493	Sequence 15493, A
C 40	47.2	4.2	2157	4	US-09-252-991A-13294	Sequence 15294, A
41	47.2	4.2	2361	4	US-09-252-991A-13546	Sequence 15546, A
42	47	4.1	891	4	US-09-252-991A-1176	Sequence 1176, App
43	47	4.1	987	4	US-09-252-991A-14271	Sequence 14271, A
44	47	4.1	996	2	US-08-924-759-21	Sequence 21, Appl
45	47	4.1	996	3	US-09-248-335-21	Sequence 21, Appl

## ALIGNMENTS

RESULT 1  
US-09-328-111-215/c  
; Sequence 215, Application US/09328111  
; Patent No. 626333  
; GENERAL INFORMATION:  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie  
; APPLICANT: Carino, Theodore J.  
; APPLICANT: Derti, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; TITLE OF INVENTION: PRODUCTS  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/328,111  
; NUMBER OF SEQ ID NOS: 850  
; EARLIER FILING DATE: 1999-06-08  
; EARLIER APPLICATION NUMBER: US 60/088,801  
; EARLIER FILING DATE: 1998-06-10  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 215  
; LENGTH: 618  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(618)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-328-111-215

Query Match	11.1%	Score 126;	DB 3;	Length 618;
Best Local Similarity	85.9%	Pred. No. 1.4e-21;		
Matches 165;	Conservative	0;	Mismatches 24;	Indels 3;
Gaps	3;			
QY	945	CCAGTACCACAGGACACTTGGGGCCCTCTCGGATGAGCACTTCCAGCCCAAGCTGAAGC	1004	
DB	534	CNAGTCCCNAGNCCANTGGGCGCTTNGAATGAGCANTTCCAGCCCAAGCTTAAGC	475	
QY	1005	AGTTCACTTCTTAATGTGACCA-CACTCCAGAGAGGC-TGTATGACTTT-CCTAAGGA	1061	
DB	474	AGTTCACTTCTTAATGTGNCACCACTCCAGANGAAGCTTGTATGACTTTGNTAAGGA	415	
QY	1062	AAATATATGATGATGATGAGGAGAAAGTTGTGGAAATATGTGATGACCTTTGGAAC	1121	

Db 414 AAATATATGATGATGAGGAGAGATTGCTGGAAATATGTGGATGACCTCTTGGAACT 355

QY 1122 GGAGGAGACAG 1133

Db 354 GGAGGAGACAG 343

## RESULT 2

US-08-125-468-1/c  
; Sequence 1, Application US/08125468  
; Patent No. 559385  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, Michael J.  
; APPLICANT: Lotvin, Jason A.  
; APPLICANT: Strathy, Nancy  
; APPLICANT: Fantini, Susan E.  
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
; TITLE OF INVENTION: chlortetracycline and tetracycline formation and cosmid  
; TITLE OF INVENTION: useful therein  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: American Cyanamid Company  
; STREET: One Cyanamid Plaza  
; CITY: Wayne  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07470  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/125,468  
; FILING DATE: 22-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tsevdos, Estelle J.  
; REGISTRATION NUMBER: 31,145  
; REFERENCE/DOCKET NUMBER: 31,255-02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201)831-3241  
; TELEFAX: (201)831-3305  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30001 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-125-468-1

Query Match 4.8%; Score 54.4; DB 1; Length 30001;  
Best Local Similarity 48.9%; Pred. No. 0.0015;  
Matches 173; Conservative 1; Mismatches 177; Indels 3; Gaps 1;  
QY 40 CCACGGAGAGCTGATGCGCTCGAGTCGGCTACCGGACGCGCTGGACAGTACAGC 99  
Db 6200 CCGGGCCGACTCCGTGCTGCTACCTCCGACGCGGCGGCGCACCGAGGACC 6141  
QY 100 GCGAGACTGGCCGAGAGCSTKGGCTACCTGGAGATCAGCTGCGGTGACCGCTTG 159  
Db 6140 CGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6081  
QY 160 CTGCGGACAGGAGGCTTCTCCACCGCACTGCGGCGCGCGCGCGCGCGCGCGCC 219  
Db 6080 ATGCGCTGATCGGCGACTCCACGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6021  
QY 220 GCGCGCGGCTCGCGAGCTATCCGAGCTCGGCTTCTCGGGGGCTGTGTGCGCGCGG 279  
Db 6020 GACCTCCGCGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5961  
QY 280 CACTGCTCAAGCGCTGCAAGAGGCGCTTCCGACGCTTCCGACAGTCCCGAGCCGC 339

Db 5960 CCGTCCCGGAGCCGCAAGTGGCGCTGTGTGGACACGGTGACACCGAA---CGCGCC 5904

QY 340 GAGGTCTGCGGACTTCCAGCGCGCGGAGCCCTACAAGTTCTCTGAGTTGCT 393

Db 5903 GAGGACCGGAGCTGTGTCGCGCGCGCTGCGCGGACAGATCGGCGCGGCGCT 5850

## RESULT 3

US-08-474-933-1/c  
; Sequence 1, Application US/08474933  
; Patent No. 586410  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, Michael J.  
; APPLICANT: Lotvin, Jason A.  
; APPLICANT: Strathy, Nancy  
; APPLICANT: Fantini, Susan E.  
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
; TITLE OF INVENTION: chlortetracycline and tetracycline formation and cosmid  
; TITLE OF INVENTION: useful therein  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: American Cyanamid Company  
; STREET: One Cyanamid Plaza  
; CITY: Wayne  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07470  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/474,933  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/125,468  
; FILING DATE: 22-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tsevdos, Estelle J.  
; REGISTRATION NUMBER: 31,145  
; REFERENCE/DOCKET NUMBER: 31,255-02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201)831-3241  
; TELEFAX: (201)831-3305  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30001 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-474-933-1

Query Match 4.8%; Score 54.4; DB 2; Length 30001;  
Best Local Similarity 48.9%; Pred. No. 0.0015;  
Matches 173; Conservative 1; Mismatches 177; Indels 3; Gaps 1;  
QY 40 CCACGGAGAGCTGATGCGCTCGAGTCGGCTACCGGACGCGCTGGACAGTACAGC 99  
Db 6200 CCGGGCCGACTCCGTGCTGCTACCTCCGACGCGGCGGCGCACCGAGGACC 6141  
QY 100 GCGAGACTGGCCGAGAGCSTKGGCTACCTGGAGATCAGCTGCGGTGACCGCTTG 159  
Db 6140 CGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6081  
QY 160 CTGCGGACAGGAGGCTTCTGCCACCGCACTGCGGCGCGCGCGCGCGCGCGCC 219  
Db 6080 ATGCGCTGATCGGCGACTCCACGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6021  
QY 220 GCGCGCGGCTCGCGAGCTATCCGAGCTCGGCTTCTCGGGGGCTGTGTGCGCGCGG 279

Db 6020 GACCTCCCGGAGGACCACTGCTGACACGAGCCCGGACCCCTGCTGCTGTC 5961  
QY 280 CACTGCTCAAGCGTGTGAAGCAGGCGCTGCGAGCTTCCGCGCAGTCCAGCCCGGCGC 339  
Db 5960 CCGTTCGCCGAGCCCGGCAAGTGGGCGTGTGTGACACGCTGGACACCGA---CGGCGCC 5904  
QY 340 GAGGTGCTGGCGACTTCCAGCGCGCGGACCTTACAGTTCTTCAGTTGCT 393  
Db 5903 GAGGACCGGAGTGTGCTCCGCGCGCCCTGCGCGACAGATCGGCGCGGGCGCT 5850

RESULT 4

US-07-951-715A-6  
; Sequence 6, Application US/07951715A  
; Patent No. 5825136  
; GENERAL INFORMATION:  
; APPLICANT: Koziel, Michael G.  
; APPLICANT: Desai, Nalini M.  
; APPLICANT: Lewis, Kelly S.  
; APPLICANT: Kramer, Vance C.  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Evola, Stephen V.  
; APPLICANT: Crossland, Lyle D.  
; APPLICANT: Wright, Martha S.  
; APPLICANT: Merlin, Ellis J.  
; APPLICANT: Launis, Karen L.  
; APPLICANT: Rothstein, Steven J.  
; APPLICANT: Bowman, Cindy G.  
; APPLICANT: Dawson, John L.  
; APPLICANT: Dunder, Erik M.  
; APPLICANT: Pace, Gary M.  
; APPLICANT: Suttie, Janet L.  
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE  
; NUMBER OF SEQUENCES: 94  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
FILING DATE: 25-SEP-1992  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Sprull, W. Murray  
REGISTRATION NUMBER: 32,943  
REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8615  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3624 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS

LOCATION: 1..3621  
OTHER INFORMATION: /product= "Full-length, maize  
OTHER INFORMATION: optimized cry1B"  
OTHER INFORMATION: /note= "Disclosed in Figure 6."  
US-07-951-715A-6  
Query Match 4.6%; Score 52; DB 1; Length 3624;  
Best Local Similarity 47.5%; Pred. No. 0.0027;  
Matches 151; Conservative 1; Mismatches 166; Indels 0; Gaps 0;  
QY 100 GGGGAGCACTGGGCGGAGAGCTKGGCTACTCTGGAGATCAGCTGCGGTGACACCGCTTG 159  
Db 181 GCGAGCTGTGGCCCCCGCGCGGACCACTGGAGATCTTCTGGAGCAGCTGGAGCAG 240  
QY 160 CTGCGGACAGCGAGGCTTCTGCCACCGCAATGCGAGCGCGCGCGGAGCGGAGCGGCGC 219  
Db 241 CTGATCAACAGCAGATCAGGAGAACGCCCGCAACACCGCCCTGCGCGCGGCG 300  
QY 220 GCGCGCGCTCGCCAGCTATCCGAGCTGGCTCTTCGGGGCGCTGCTGCGCGCGCG 279  
Db 301 CTGGGCGACAGCTTCCGCGGCTACCGAGAGCTGGCTGGAGACTGGCTGGAGAACCGGAC 360  
QY 280 CACTGCTCAAGCGCTGCAAGCAGGCGCTGCCAGCTTCCGCGAGTCCCGAGCGCGCGCGC 339  
Db 361 GACGCGCGCAGCGCGAGCGTGTGTACACCCAGTACATCGCCCTGGAGCTGGACTTCTG 420  
QY 340 GAGGTGCTGGCGGACTTCCAGCGCGCGGAGCGGCTTACAGTTCTTCGAGTTGCTTACTTC 399  
Db 421 AACGCTATGCGCTGTTCGCCATCCGCAACAGAGGTGCGCCCTGCTGATGTTGCTACGCC 480  
QY 400 AAGCAATAATATCTCCCC 417  
Db 481 CAGCGCGCAACTGCGAC 498

RESULT 5

US-08-459-448A-6  
; Sequence 6, Application US/08459448A  
; Patent No. 5959336  
; GENERAL INFORMATION:  
; APPLICANT: Koziel, Michael G.  
; APPLICANT: Desai, Nalini M.  
; APPLICANT: Lewis, Kelly S.  
; APPLICANT: Kramer, Vance C.  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Evola, Stephen V.  
; APPLICANT: Crossland, Lyle D.  
; APPLICANT: Wright, Martha S.  
; APPLICANT: Merlin, Ellis J.  
; APPLICANT: Launis, Karen L.  
; APPLICANT: Rothstein, Steven J.  
; APPLICANT: Bowman, Cindy G.  
; APPLICANT: Dawson, John L.  
; APPLICANT: Dunder, Erik M.  
; APPLICANT: Pace, Gary M.  
; APPLICANT: Suttie, Janet L.  
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
; TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE  
; NUMBER OF SEQUENCES: 94  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5859336artis Corporation  
; STREET: Patent & Trademark Dept., 520 White Plains  
; CITY: Tarrytown  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10591-9005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/459,448A  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40403  
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8582  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 6:  
LENGTH: 3624 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3621  
OTHER INFORMATION: /product= "Full-length, maize  
OTHER INFORMATION: optimized cryiB"  
OTHER INFORMATION: /note= "Disclosed in Figure 6."  
US-08-459-448A-6

Query Match 4.6%; Score 52; DB 2; Length 3624;  
Best Local Similarity 47.5%; Pred. No. 0.0027;  
Matches 151; Conservative 1; Mismatches 166; Indels 0; Gaps 0;

QY 100 GCGAGAGCTGGCGGCGAGCTGGCTACCTGGAGATCAGCTCGGCTGCACGGCTG 159  
DB 181 GCGAGAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
QY 160 CTGCGCGAGAGCTGGCTGGCTACCTGGAGATCAGCTCGGCTGCACGGCTG 219  
DB 241 CTGATCAACAGCAGATACCCAGAGACGCCGCCACCGGCTGGCGGCTGCAGGGC 300  
QY 220 GCGCGCGGCTCGCAGCTATCCCGAGTGCCTCTTGGGGGCGCTGCTGGCGGCG 279  
DB 301 CTGGGCGAGCTTCGGCGCTACCGAGAGAGCTGGAGAGCTGGAGAAACCGGAC 360  
QY 280 CACTGCTCAAGGCTGAGCAGGCGCTGCGAGGCTGCGAGCTTCCGCGAGTCCCGAGCCGCGC 339  
DB 361 GACGCGCGCAGCTTCGGCGCTACCGAGTATCCCGAGTATCCCGAGTATCCCGAGTATCCCGAGT 420  
QY 340 GAGTGTGGCGAGCTTCCAGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 399  
DB 421 AAGCGCATGCCCTGTTCGCCATCCGCAACAGGAGTCCCGCTGATGTGTACGCC 480  
QY 400 AAGGCAAAATATCTCCC 417  
DB 481 CAGGCGCGCAACCTGCAC 498

RESULT 6  
US-08-459-595A-6  
Sequence 6, Application US/08459595A  
Patent No. 6018104  
GENERAL INFORMATION:  
APPLICANT: Koziel, Michael G.  
APPLICANT: Desai, Nalini M.  
APPLICANT: Lewis, Kelly S.  
APPLICANT: Kramet, Vance C.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Evola, Stephen V.

APPLICANT: Crossland, Lyle D.  
APPLICANT: Wright, Martha S.  
APPLICANT: Merlin, Ellis J.  
APPLICANT: Launis, Karen L.  
APPLICANT: Rothstein, Steven J.  
APPLICANT: Bowman, Cindy G.  
APPLICANT: Dawson, John L.  
APPLICANT: Dunder, Erik M.  
APPLICANT: Pace, Gary M.  
APPLICANT: Suttie, Janet L.  
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED  
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6018104artis Corporation  
STREET: Patent & Trademark Dept., 520 White Plains  
STREET: Rd., POB 2005  
CITY: Tarrytown  
STATE: New York  
COUNTRY: USA  
ZIP: 10591-9005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,595A  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40403  
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8582  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3624 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3621  
OTHER INFORMATION: /product= "Full-length, maize  
OTHER INFORMATION: optimized cryiB"  
OTHER INFORMATION: /note= "Disclosed in Figure 6."  
US-08-459-595A-6

Query Match 4.6%; Score 52; DB 3; Length 3624;  
Best Local Similarity 47.5%; Pred. No. 0.0027;  
Matches 151; Conservative 1; Mismatches 166; Indels 0; Gaps 0;

QY 100 GCGAGAGCTGGCGGCGAGCTGGCTACCTGGAGATCAGCTCGGCTGCACGGCTG 159  
DB 181 GCGAGAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
QY 160 CTGCGCGAGAGCTGGCTGGCTACCTGGAGATCAGCTCGGCTGCACGGCTG 219  
DB 241 CTGATCAACAGCAGATACCCAGAGACGCCGCCACCGGCTGGCGGCTGCAGGGC 300  
QY 220 GCGCGCGGCTCGCAGCTATCCCGAGTGCCTCTTGGGGGCGCTGCTGGCGGCG 279

Db 301 CTGGGCGACAGCTTCCGGCGCTACACAGCAGAGCTGGAGGACTGGCTGGAGAACCGGCAC 360  
QY 280 CACTGCCTCAAGCGCTGAACAGGGGCTGCAGCCTTCCGCCAGTCCAGCCAGCGGC 339  
Db 361 GACCCCGCACCAGCGTCTGTACACCGAGTACATCGCCCTGGAGCTGGACTTCTTG 420  
QY 340 GAGGTGCTGGCGGACTTCCAGCGCGGAGCGCTTACAGTTCTCTGCAAGTTCTGCTTACTTC 399  
Db 421 AACGCCATGCCCTTTCGCCATCGCAACAGAGGTGCCCTGCTGATGGTGTACGCC 480  
QY 400 AAGGCAATAATCTCCCC 417  
Db 481 CAGGCGCGCAACCTGCAC 498

RESULT 7

US-08-459-504B-6

; Sequence 6, Application US/08459504B

; Patent No. 6075185

; GENERAL INFORMATION:

; APPLICANT: Koziel, Michael G.

; APPLICANT: Desai, Nalini M.

; APPLICANT: Lewis, Kelly S.

; APPLICANT: Kramer, Vance C.

; APPLICANT: Warren, Gregory W.

; APPLICANT: Evola, Stephen V.

; APPLICANT: Crossland, Lyle D.

; APPLICANT: Wright, Martha S.

; APPLICANT: Merlin, Ellis J.

; APPLICANT: Launis, Karen L.

; APPLICANT: Rothstein, Steven J.

; APPLICANT: Bowman, Cindy G.

; APPLICANT: Dawson, John L.

; APPLICANT: Dunder, Erik M.

; APPLICANT: Pace, Gary M.

; APPLICANT: Suttle, Janet L.

; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED

; INSECTICIDAL ACTIVITY IN MAIZE

; NUMBER OF SEQUENCES: 94

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6075185artis Corporation

; STREET: 3054 Cornwallis Road

; CITY: Research Triangle Park

; STATE: NC

; COUNTRY: USA

; ZIP: 27709

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/459,504B

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/459,595

; FILING DATE: 02-JUN-1995

; APPLICATION NUMBER: US 07/951,715

; FILING DATE: 25-SEP-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/772,027

; FILING DATE: 04-OCT-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Meigs, J. Timothy

; REGISTRATION NUMBER: 38,241

; REFERENCE/DOCKET NUMBER: CGC1577/CIP/DIV

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (919) 541-8587

; TELEFAX: (919) 541-8689

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3624 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "Synthetic DNA"  
; HYPOTHETICAL: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3621  
; OTHER INFORMATION: /product= "Full-length, maize  
; OTHER INFORMATION: Optimized cryiB"  
; OTHER INFORMATION: /note= "Disclosed in Figure 6."  
; US-08-459-504B-6

Query Match 4.6%; Score 52; DB 3; Length 3624;

Rest Local Similarity 47.5%; Pred. No. 0.0027;

Matches 151; Conservative 1; Mismatches 166; Indels 0; Gaps 0;

QY 100 GCGAGCACTGGCGGAGAGCTKGGCTACCTGGAGATCAGCCTGGGGTGCACCGCTTG 159  
Db 181 GCGAGCTGTGGCCCGCGCGCGACCACTGGAGATCTTCTGGAGCAGCTGGAGCAG 240  
QY 160 CTGGCGGACAGGAGGCTTCTGCCACCGCACTCAGCGCGCGCGAGCCGAGCCG 219  
Db 241 CTGATCAACAGCAGATCACCGAGAACGCCCGCAACACCGCCCTGGCCGCTGCAGGC 300  
QY 220 GCGCGGCGCTCGCCAGCTATCCCGAGCTGCGCTTCTGGGGGCTGCTGGCGCGCG 279  
Db 301 CTGGGCGACAGCTTCGCGCTTACCAGCAGAGCTGGAGGACTGGCTGGAGAACCG 360  
QY 280 CACTGCTCAAGGCTGCAAGCAGGCGCTGCCAGCTTCCGCGAGTCCCGAGCCGAG 339  
Db 361 GACGCCCGCAGCCGAGCGCTGTGTACACCCAGTATCGCCCTGGAGTCTCTG 420  
QY 340 GAGGTGCTGGCGGACTTCCAGCGCGCGAGCGCTACAGTTCCTGAGTTCGCTTACTTC 399  
Db 421 AAGCCATGCCCTGTTCGCCATCGCAACAGGAGTGGCCCTGCTGATGTGTACGCC 480  
QY 400 AAGGCAATAATCTCCCC 417  
Db 481 CAGGCGCGCAACCTGCAC 498

RESULT 8

US-08-459-444-6

; Sequence 6, Application US/08459444A

; Patent No. 6121014

; GENERAL INFORMATION:

; APPLICANT: Koziel, Michael G.

; Desai, Nalini M.

; Lewis, Kelly S.

; Kramer, Vance C.

; Warren, Gregory W.

; Evola, Stephen V.

; Crossland, Lyle D.

; Wright, Martha S.

; Merlin, Ellis J.

; Launis, Karen L.

; TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED

; NUCLEIC ACID CODING SEQUENCE

; NUMBER OF SEQUENCES: 94

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6121014artis Agribusiness Biotechnology Research, Inc.

; STREET: 3054 Cornwallis Road

; CITY: Research Triangle Park

; STATE: NC

; COUNTRY: USA

; ZIP: 27709

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,444A  
FILING DATE: 02-Jun-1995  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/951,715  
FILING DATE: 25-SEP-1992  
APPLICATION NUMBER: US 07/772,027  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Weigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-18805/P1/CGC1577/CIP/DIV6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8587  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3624 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3621  
OTHER INFORMATION: /product= "Full-length, maize  
optimized cryiB"  
/note= "Disclosed in Figure 6."  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-08-459-444-6  
Query Match 4.6%; Score 52; DB 3; Length 3624;  
Best Local Similarity 47.5%; Pred. No. 0.0027;  
Matches 151; Conservative 1; Mismatches 166; Indels 0; Gaps 0;  
QY 100 GCGAGCAGCTGGCGCGAGAGCSTKGCTACCTGGAGATCAGCTCGCGCTGCACCGCTTG 159  
DB 181 GCGAGCTGTGGCGCGCGCGCGAGCAGTGGAGATCTTCTGGAGCAGCTGGAGCAG 240  
QY 160 CTGGCGGAGAGGAGCCTTTGCGACCGCACTGCGAGCGCGCGCGAGCGCCGAGCGCC 219  
DB 241 CTGATCAACAGCAGATCACCAGAACCGCCCGCAACCGCCCTGGCGCGCTGGAGGGC 300  
QY 220 GCGCGCGCTGCGCAGCTATCCGAGCTGCGCTCTTGGGGGGCTGTGCGCGCGCG 279  
DB 301 CTGGGCGAGAGCTTCGCGCCTACAGCAGAGCCTGGAGACTGGCTGGAGAACCGCGAC 360  
QY 280 CACTGCTCAAGCGGTGCAAGAGGCGCTTCCAGCGCTTCCGCGAGTCCCGAGCCAGCGCGC 339  
DB 361 GAGCGCGGACCGCGAGCGTGTGTACACCCAGTACATCGCCCTGGAGCTGACTTCCTG 420  
QY 340 GAGTCTGCGGAGCTTCCAGCGCGCGAGCCCTACAAGTTCTCTGAGTTCTGCTTACTTC 399  
DB 421 AAGCGATGCCCTGTTCGCGATCCGCAACAGGAGTGGCCCTGCTGATGTTGATGCC 480  
QY 400 AAGCGAAATAATCTCCCC 417  
DB 481 CAGCGCGCAACCTGCAC 498  
RESULT 9  
US-09-553-549-7  
; Sequence 7, Application US/09053549  
; Patent No. 6121521  
; GENERAL INFORMATION:  
; APPLICANT: Desai, Nalini  
; TITLE OF INVENTION: No. 6121521el Insecticidal Protein and Gene  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
ADDRESS: No. 6121521artis Corporation  
STREET: 3054 Cordwallis Rd.  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/053,549  
FILING DATE: 01-APR-1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1995  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3624 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3621  
OTHER INFORMATION: /product= "Full-length, maize  
optimized cryiB"  
OTHER INFORMATION: /note= "Disclosed in Figure 6."  
US-09-053-549-7  
Query Match 4.6%; Score 52; DB 3; Length 3624;  
Best Local Similarity 47.5%; Pred. No. 0.0027;  
Matches 151; Conservative 1; Mismatches 166; Indels 0; Gaps 0;  
QY 100 GCGAGCAGCTGGCGCGAGAGCSTKGCTACCTGGAGATCAGCTCGCGCTGCACCGCTTG 159  
DB 181 GCGAGCTGTGGCGCGCGCGCGAGCAGTGGAGATCTTCTGGAGCAGCTGGAGCAG 240  
QY 160 CTGGCGGAGAGGAGCCTTTGCGACCGCACTGCGAGCGCGCGCGAGCGCCGAGCGCC 219  
DB 241 CTGATCAACAGCAGATCACCAGAACCGCCCGCAACCGCCCTGGCGCGCTGGAGGGC 300  
QY 220 GCGCGCGCTGCGCAGCTATCCGAGCTGCGCTCTTGGGGGGCTGTGCGCGCGCG 279  
DB 301 CTGGGCGAGAGCTTCGCGCCTACAGCAGAGCCTGGAGACTGGCTGGAGAACCGCGAC 360  
QY 280 CACTGCTCAAGCGGTGCAAGAGGCGCTTCCAGCGCTTCCGCGAGTCCCGAGCCAGCGCGC 339  
DB 361 GAGCGCGGACCGCGAGCGTGTGTACACCCAGTACATCGCCCTGGAGCTGACTTCCTG 420  
QY 340 GAGTCTGCGGAGCTTCCAGCGCGCGAGCCCTACAAGTTCTCTGAGTTCTGCTTACTTC 399  
DB 421 AAGCGATGCCCTGTTCGCGATCCGCAACAGGAGTGGCCCTGCTGATGTTGATGCC 480  
QY 400 AAGCGAAATAATCTCCCC 417  
DB 481 CAGCGCGCAACCTGCAC 498  
RESULT 10  
US-09-547-422-6  
; Sequence 6, Application US/09547422  
; Patent No. 6320100  
; GENERAL INFORMATION:

APPLICANT: Kozziel, Michael G.  
Desai, Nalini M.  
Lewis, Kelly S.  
Kramer, Vance C.  
Warren, Gregory W.  
Evola, Stephen V.  
Crossland, Lyle D.  
Wright, Martha S.  
Merlin, Ellis J.  
Lauais, Karen L.

TITLE OF INVENTION: SYNTHETIC  
INSECTICIDE

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 6320100ar  
STREET: 3054 Cornwallis  
CITY: Research Triangle  
STATE: NC  
COUNTRY: USA

## RESULT 11

```

US-09-252-991A-3976
; Sequence 3976, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO
; ACID SEQUENCES FOR THE DIAGNOSIS OF
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3976
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3976

```

## RESULT 12

US-09-252-991A-4046 ; Sequence 4046, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:



APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4046  
; LENGTH: 1521  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4046

Query Match 4.5%; Score 50.6; DB 4; Length 1521;  
Best Local Similarity 46.4%; Pred. No. 0.0042; Mismatches 185; Indels 0; Gaps 0;  
Matches 161; Conservative 1;  
QY 26 GTTCCGAGCTTCCACGGGACGAGCTGATGCGCTCGAGTCGGCCTACCGGCAAGCGC 85  
DB 189 GCATCGCATCTTACCCCGGTCAACGAACTGCGGTTCGCGGCCACCGCTGCTGGCA 248  
QY 86 TGGACAAGTACAGCGGCGAGCACTGGGCGGAGAGCTGGGCTACTTGGAGATCAGCCTGC 145  
DB 249 CGCCATCCCTCGGCGCGGACGACGACAGGAGGCTGTTCCTGGAGACGAGATGG 308  
QY 146 GGCTGACCGCTGCTGCGCGACGAGCGAGGCTTCTGCCACCGCAACTGCAGCGCGCGC 205  
DB 309 GCACCGTGCCCTTGGCCCTGGAGCGGACGAGCAAGGTGTGGCTTGCAGTCAGCAGC 368  
QY 206 CGGAGCCCGAGCGCGCGCGCTCGCCAGCTATCCGAGTATCCGAGTGTGCGCTTTCGGGGGCC 265  
DB 369 AGCCGATCCCGACCTGGGACATTCAGCGCGCGCGGAGCTGTGCGCGCCCTCGGCC 428  
QY 266 TGTGGCGCGCGGCACTGCTCAGCGCTGCAAGCAGGCGCTGCGAGGCTTCCGCCAGT 325  
DB 429 TGAAGGCTCGACCTTTCCTCATCGAGTCTACGCAACGGCGCGCGGCACTTTCGTGCG 488  
QY 326 CCAGCCCGAGCGCGAGGTGCTGGCGACTTCCAGCGCGCGAGGCC 372  
DB 489 GCCTGGAGAGCGTGGCGCGCTGTCGCGCTGCACCGCCAGCACCGC 535

RESULT 13  
US-09-105-537-30  
; Sequence 30, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30  
; LENGTH: 13842  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-30

Query Match 4.4%; Score 49.8; DB 3; Length 13842;  
Best Local Similarity 48.4%; Pred. No. 0.015; Mismatches 147; Indels 0; Gaps 0;  
Matches 138; Conservative 0;  
QY 108 CTGGGCGAGAGCGTGGCTTACCTGGAGATCAGCTTCCGCTGCGCTTGTGCGGCA 167

DB 9519 CCGGAGCGGCACACCGCGTCTCTTACCGGCCAGGCGAGTCAGCGGCCCGCGCGCG 9578  
QY 168 CAGGAGCGCTTTCGCCACCGCAACTGACGCGCGCGCGCGCGCGCGCGCGCGCG 227  
DB 9579 CCGTGAAGTACAGACCGGACCCCGCTTTCGCCCGGGGCTCGACGAGATCTGCGCCCA 9638  
QY 228 CTTGCGCAGCTATCCGAGAGCTGCGCTTTCGGGGGCGCTGCTGCGCGCGCGCGCGCT 287  
DB 9639 CTTGCGCGGTCACTCGAACTGCGCTTTCGAGACGTGATGTTTCGGCGCGCGAGGCGCG 9698  
QY 288 CAAGCGCTCAAGCAGGCGCTGCGAGCTTTCGCCAGTCCCGAGTCCCGCGCGCGCGCGCT 347  
DB 9699 GAGGCGCGCTGCTGACGAGACGCGGTACACGAGTGGCGCTTTCGCCCTGAGGT 9758  
QY 348 GCGGAGCTTTCAGCGCGCGCGAGCGCTTTCAGAGTTCCTGCGAGTTCGC 392  
DB 9759 CGCGCTTTCGCGCTGTCGAGAGCTGGGCGCATGCGCGCGCGCGCG 9803

RESULT 14  
US-09-105-537-5  
; Sequence 5, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 36778  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-5

Query Match 4.4%; Score 49.8; DB 3; Length 36778;  
Best Local Similarity 48.4%; Pred. No. 0.021; Mismatches 147; Indels 0; Gaps 0;  
Matches 138; Conservative 0;  
QY 108 CTGGGCGAGAGCGTGGCTTACCTGGAGATCAGCTTTCGCGCGCGCGCGCGCGCG 167  
DB 11260 CCGGAGCGGCACACCGCTTCTTTCACCGCGCGAGGCGAGTCAGCGCGCGCGCGCG 11319  
QY 168 CAGGAGCGCTTTCGCCACCGCAACTGACGCGCGCGCGCGCGCGCGCGCGCGCG 227  
DB 11320 CCGTGAAGTACAGACCGGACCCCGCTTTCGCCCGGGGCTCGACGAGATCTGCGCCCA 11379  
QY 228 CTTGCGCAGCTATCCGAGAGCTGCGCTTTCGGGGGCGCTGCTGCGCGCGCGCGCT 287  
DB 11380 CTTGCGCGGTCACTCGAACTGCGCTTTCGAGTGTGCTGAGTGTGCTGCGCGCGCGCGCG 11439  
QY 288 CAAGCGCTCAAGCAGGCGCTGCGAGCTTTCGCCAGTCCCGAGTCCCGCGCGCGCGAGTGT 347  
DB 11440 GAGGCGCGCTGCTGACGAGACGCGGTACACGAGTGGCGCTTTCGCCCTGAGGT 11499  
QY 348 GCGGAGCTTTCAGGCGCGCGAGCGCTTTCAGAGTTCCTGCGAGTTCGC 392  
DB 11500 CGCGCTTTCGCGCTGTCGAGAGCTGGGCGCATGCGCGCGCGCGCG 11544

RESULT 15  
US-09-320-878-19  
; Sequence 19, Application US/09320878A  
; Patent No. 6117659  
; GENERAL INFORMATION:  
; APPLICANT: Ashley, Gary  
; APPLICANT: Betlach, Melanie C.  
; APPLICANT: Betlach, Mary C.

```
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONCLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match      4.4%; Score 49.8; DB 3; Length 38506;
Best Local Similarity 48.4%; Pred. No. 0.022;
Matches 138; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 108 CTGGGCGGAGAGCTGGCTACCTGGAGATCAGCTCGGCTGCACCGCTTGTCTGGCGGA 167
Db 9402 CCGGACGGACACACCGCGTCTTTCACCGGCGAGGAGTCAAGCCCGCGCGCG 9461

QY 168 CAGCGAGGCTTCTGCCACCGCAACTCGAGCGCGCGCGCGCGAGCCCGCGCGCG 227
Db 9462 CCGTGAGCTGTACGACCGGACCCCGCTTTCGCCCGGCGCTCGACGAGATCTGCGCCCA 9521

QY 228 CCTCGCCAGCTATCCGAGCTGGCTCTTTCGGGGGCGCTCTCGCGCGCGCGCCT 287
Db 9522 CCTCGAGGCTACCTCGAACTGCCCTGCTCGAGCTGATGTTTCGGCGCGAGGCGAGCG 9581

QY 288 CAAGCGGTGCAAGCAGGCGCTGCGAGCTTTCGCCAGTCCCGAGCCCGAGCGGTGCT 347
Db 9582 GGAGGCGCGCTGTCTGACGAGCGCGGTACAGCAGTGGCGCTGTTCCGCTTGAGGT 9641

QY 348 GGCGGACTTCCAGCGCGCGAGCCCTACAGTTCCTGCACTTCCG 392
Db 9642 CGCGCTTTCGGGCTCGTCGAGAGCTGGGGCATCGCGCGCGCGC 9686
```

Search completed: January 30, 2004, 18:32:30  
Job time : 69.9336 secs





```
Db 191 GGCCTACCGGACGCGCTGGACAGTACAGCGCGGAGCAGCTGKCGGAGAGCGTGGGCTA 250
Qy 129 CTGAGATAGCTGCGGCTGCACCGCTTCTGCGGACAGCGAGCGCTTCTGCCACCG 188
Db 251 CCTGAGATAGCTGCGGCTGCACCGCTTCTGCGGACAGC-AGGCTTCTGCCACCG 309
Qy 189 CAATGACGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 248
Db 310 CAATGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 369
Qy 249 GCGCTCTTCTGGGCGGCTGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 308
Db 370 GCGCTCTTCTGGGCGGCTGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 429
Qy 309 GCGAGCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 368
Db 430 GCGAGCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 489
Qy 369 GCGCTACAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 428
Db 490 GCGCTACAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 549
Qy 429 GCGCTACAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 488
Db 550 GCGCTACAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 609
Qy 489 ATATTATAAGAGCGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 548
Db 610 ATATTATAAGAGCGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 669
Qy 549 TGAAGCGCTTCTATCGAGAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 608
Db 670 TGAAGCGCTTCTATCGAGAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 729
Qy 609 CACAGATAGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 668
Db 730 CACAGATAGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 789
Qy 669 CTGCGAGGTTCCAGGAGATCAAGGACTTCAAGGACTTCAAGGACTTCAAGGACTTCAAGGACTTCA 728
Db 790 CTGCGAGGTTCCAGGAGATCAAGGACTTCAAGGACTTCAAGGACTTCAAGGACTTCAAGGACTTCA 849
Qy 729 TTATGTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAGAACTCAAGGACTTCAAGGACTTCAAGG 788
Db 850 TTATGTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAGAACTCAAGGACTTCAAGGACTTCAAGG 909
Qy 789 AGGCTATCGGTTGAGAAATTTCTGCTACCATGATATCAATTAATCTGAGTGTGCTGCTATTA 848
Db 910 AGGCTATCGGTTGAGAAATTTCTGCTACCATGATATCAATTAATCTGAGTGTGCTGCTATTA 969
Qy 849 TAAGTTGAACGACCTGAAGATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 908
Db 970 TAAGTTGAACGACCTGAAGATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1029
Qy 909 GAATGAACAGGTCATGACGAGAACTGCTGTGATTAACAGTGTGAAGAGAACTCAAGGACTTCAAGG 968
Db 1030 GAATGAACAGGTCATGACGAGAACTGCTGTGATTAACAGTGTGAAGAGAACTCAAGGACTTCAAGG 1089
Qy 969 CTTG 972
Db 1089 YVVC 1092

RESULT 3
LOCUS AL514488
DEFINITION Homo sapiens NEUROBLASTOMA 1201 bp mRNA linear EST 08-MAY-2003
ACCESSION CLOBB003ZF09 5-PRIME, mRNA sequence.
VERSION AL514488
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
On Feb 13, 2001 this sequence version replaced gi:12777992.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 2233.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CLOBB003ZF09RPI&cluster=2233.f. Contact :  
Peng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CLOBB003ZF09RPI.  
Location/Qualifiers  
1..1201

FEATURES  
source

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CLOBB003ZF09"  
/tissue\_type="NEUROBLASTOMA"  
/clone\_lib="Homo sapiens NEUROBLASTOMA"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."  
BASE COUNT 274 a 299 c 324 g 246 t 58 others  
ORIGIN

Query Match 80.1%; Score 908.2; DB 9; Length 1201;  
Best Local Similarity 93.7%; Pred. No. 8.2e-210;  
Matches 953; Conservative 31; Mismatches 29; Indels 4; Gaps 4;

```
Qy 2 GCGCGGCCAATACGACGCTACAGCTTCGCGAGCTTCCACGGGACGAGTGTATGCCG 61
Db 132 GCGCGGCCAATACGACGCTACAGCTTCGCGAGCTTCCACGGGACGAGTGTATGCCG 191
Qy 62 TCGAGTCGCGCTACCGGACGCGCTGACAGTACAGGACGAGGCGGAGCTGGCCGAGGCS 121
Db 192 TCGAGTCGCGCTACCGGACGCGCTGACAGTACAGGACGAGGCGGAGCTGGCCGAGGCG 251
Qy 122 TGGCTACCTGGAGATCAGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCT 181
Db 252 TGGCTACCTGGAGATCAGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCT 311
Qy 182 GCCACCGCAACTGACGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTATC 241
Db 312 GCTACCGCAACTGACGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTATY 371
Qy 242 CCGAGCTGCGGCTCTTTCGGGGGCGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 301
Db 372 CCGAGCTGCGGCTCTTTCGGGGGCGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 429
Qy 302 AGGCTTGGCAGCGCTTCCGCGAGTCCCGAGCCCGAGCGGCGGAGGTCTGGCGGAGCTCCAGC 361
Db 430 AGGCTTGGCAGCGCTTCCGCGAGTCCCGAGCGGCGGAGGTCTGGCGGAGCTTTHAGC 489
Qy 362 GCGCGAGCGCTACAGTTCCTGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 421
Db 490 GCGCGAGCGCTTAAATTCCTGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 549
Qy 422 CCAATGCGCGCTGCTCAACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 481
Db 550 CCAATGCGCGCTGCTCAACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 609
Qy 482 ACATGGCATATTATAGAGCGCTGCGGCGGAGCTACATTAAGACCTTGGAACCA 541
```

```

Db      610 ACATGNNATATTATAARACATCGTGGTCTGCTGAGGACTACATTAAAGACCTGGAAACCA 669
QY      542 AGTCATATGAACCCCTTTTCATCCGAGCAGTGGCGCATACACCGTGAGAACTGGAGAA 601
Db      670 AGTTATATGAACGCTGTTTCATCCGAGCAGTGGCGCATACACCGTGAGAACTGGAGAA 729
QY      602 CATCCATCAGACATGAGAGTGGCCCTTCCCGACTTCTTCAAGCCTTTTACGAGTGTC 661
Db      730 CATCCATCAGACATGAGAGTGGCCCTTCCCGACTTCTTCAAGCCTTTTACGAGTGTC 789
QY      662 TCACAGCCTCGAGGGTTCACAGGAGATCAAGGACTTCAAGGATTTTACCTTTCCATAG 721
Db      790 TCACAGCCTCGAGGGTTCACAGGAGATCAAGGACTTCAAGGATTTTACCTTTCCATAG 849
QY      722 CAGATCATATGATGAGAGTTCGGAATGCAAAATACAGTGTGAAGAGAACCTCACCCAG 781
Db      850 CAGATCATATGATGAGAGTTCGGAATGCAAAATACAGTGTGAAGAGAACCTCACCCAG 909
QY      782 TTATAGAGGCTATCCGGTGTGAGAAATTTTGGCTACCATGATCATATCTTTCAGTCT 901
Db      910 TTATAGAGGCTATCCGGTGTGAGAAATTTTGGCTACCATGATCATATCTTTCAGTCT 969
QY      842 CTTATATAGTTGAACGACTGAGAGTGCAGCCCTTGCAGTCTTTCAGTCTTTCAGTCT 901
Db      970 CTTATATAGTTGAACGACTGAGAGTGCAGCCCTTGCAGTCTTTCAGTCTTTCAGTCT 1028
QY      902 TTGATCAGATGACAGAGTGCAGAGCACTGAGAGTGCAGCCCTTGCAGTCTTTCAGTCT 961
Db      1029 TTGATCAGATGACAGAGTGCAGAGCACTGAGAGTGCAGCCCTTGCAGTCTTTCAGTCT 1088
QY      962 CTTGGGGCTCTCGGATGAGCACTTCCAGCCAGACCTGAGAGTGCAGTCTTTCAGTCT 1018
Db      1089 CTTGGGGCTCTCGGATGAGCACTTCCAGCCAGACCTGAGAGTGCAGTCTTTCAGTCT 1144

RESULT 4
LOCUS   AK047506
DEFINITION Mus musculus 1656 bp mRNA linear HTC 05-DEC-2002
          enriched library, clone:B930079M13 product:cartilage associated
          protein, full insert sequence.
ACCESSION AK047506
VERSION   AK047506.1 GI:26339809
KEYWORDS HTC; CAP trapper.
SOURCE  Mus musculus (house mouse)
ORGANISM Mus musculus
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
          Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
          Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new genes
          Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 99279253
PUBMED   10349636
REFERENCE 2
AUTHORS Carninci, P., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P.,
          Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M.,
          Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hazada, A.,
          Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
          Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
          Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
          Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
          RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer
          Genome Res. 10 (11), 1757-1771 (2000)
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P.,
          Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M.,
          Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hazada, A.,
          Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
          Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
          Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
          Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
          RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer
          Genome Res. 10 (11), 1757-1771 (2000)

```

# MEDLINE PUBMED REFERENCE AUTHORS

20530913  
11076861  
4  
Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I.,  
Saito, T., Okazaki, Y., Gojibori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,  
Quackenbush, J., Schriml, L.M., Stauble, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, P., Bojunga, N.,  
Carninci, P., de Bernaldo, M.F., Brownstein, M.J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
Toyo-oka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,  
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.,  
and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6921), 685-690 (2001)  
21085660  
11217851  
5

The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 1656)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kuida, M.,  
Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers

# FEATURES source

1. 1656  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6J"  
/db\_xref="PANTOM,DB:B930079M13"  
/db\_xref="taxon:10090"  
/clone="B930079M13"  
/tissue type="cerebellum"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days neonate"  
70. 1272

CDS

```
/note="unnamed protein product; cartilage associated
protein (MGI:1891221, GB|NM_019922, evidence: BLASTN,
99%, match=1655)
putative"
/codon_start=1
/protein_id="BAC33076.1"
/db_xref="GI:26338810"
/translation="MGRPSPTAALLVLLVCVCAPTGPGVYRSFRSPRDELMPLE
SAVRLHQISGHWAEVGLVLSRLHRLRDSFAFCHRNCSAATPAPAGPASH
AEURLFQSVLRACQKRCQKQGLPAPFQSRQSLADFOQEPYKFLQFAYFKANDL
PKAIAAHTYLLKHFDDEMMKRMEEYKSLPQAEHDIKDLETKSYESLFFVRAVNG
ENWRTSISDEMLAPDLKAFYELCLACESGREIKDFKDFYLSIAHYHVEFLCKIRC
EETLTPVIGQVPEKPVATWYHQLQPAYKLNKNAAPCAVSLLFDOSDRWQONL
VYQYHEDKXWGLSDEHFQPRPEAVQFENWTLQKEIYDFAQELHMDDEGEVVEYVD
LLETEESA"
BASE COUNT 341 a 498 c 448 g 369 t
ORIGIN
Query Match 79.1%; Score 897.2; DB 11; Length 1656;
Best Local Similarity 87.0%; Pred. No. 4.6e-207;
Matches 983; Conservative 2; Mismatches 145; Indels 0; Gaps 0;
QY 1 GGGCGGCCCAATACGAACGCTACAGCTTCGGAGCTTCCACGGGACGAGCTGATGCCG 60
DB 136 GGGCGCGGGAGTACGAGCGCTACAGCTTCGGAGCTTCCGCGGAGGAGCTGATGCCG 195
QY 61 CTGAGTCCGCTACCGGACGCGCTGCGACAGTACAGCGGCGGAGCACTGGCGCGAGAC 120
DB 196 CTGAGTCCGCTACCGGACGCGCTGCGACAGTACAGCGGCGGAGCACTGGCGCGAGAC 255
QY 121 STKGCTACCTGGAGATCAGCCTCGGCTGACCGCTTGTCTGCGCGACAGCGGCTTC 180
DB 256 GTGGCTACCTGGAGTACGCTCGGCTGACCGCTGCTGCTGCGCGACAGCGGCTTC 315
QY 181 TGCACCGCAACTGCAGCGCGCGCGCGAGCCCGCGCGCGCGCGCGCGCGCGCGCG 240
DB 316 TGCACCGCAACTGCAGCGCGCGCGCGAGCCCGCGCGCGCGCGCGCGCGCGCGCG 375
QY 241 CCGGAGTCCGCTCTTCGGGGGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 300
DB 376 GCGGAATACGCTCTTCGGAGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 435
QY 301 CAGGCTCCGAGCTTCGCCAGTCCAGCCAGCGCGGAGGTGTGCGGAGCTTCAG 360
DB 436 CAGGCTCCGCGCTTCGCCAGTCCAGCCAGCGCGCGCTGCTGCGGAGCTTCAG 495
QY 361 CGCGGAGCGCTACAGTCTTCGAGTTCGCTTACGAGCGCAATATCTCCGCCAA 420
DB 496 CAGGCGGAGCGCTACAGTCTTCGAGTTCGCTTACGAGCGCAATATCTCCGCCAA 555
QY 421 GCATCGCGCTGCTCACACCTTCTACTGAAGCATCTCTGATGACGAAATGATGAAGG 480
DB 556 GCATCGCTCGGCTCACACCTATCTCTGAAGCATCTCTGATGACGAAATGATGAAGG 515
QY 481 AACATGCGATATTATAGAGCTGCTGTCGCGGAGCTACATTTAAAGACCTGGAACC 540
DB 616 AACATGAGTATTATAGAGCTGCTGTCGCGGAGCTACATTTAAAGACCTGGAACC 675
QY 541 AAGTCATATGAAGCGCTGTTTCATCCGAGGAGTCCGGGCTACAAACGCTGAGAACTGGAG 600
DB 676 AAGTCATGAGAGCGCTGTTTCATCCGAGGAGTCCGGGCTACAAACGCTGAGAACTGGAG 735
QY 601 ACATCCATCACAGATGAGCTGGCGCTTCGCGATCTTCTTCAAGCGCTTTTACAGTGT 660
DB 736 ACCTCCATTTCCGACATGAGCTGCGGCTTCGCGATCTTCTTCAAGCGCTTTTACAGTGT 795
QY 661 CTGCGAGCTTCGAGGCTTCGAGGAGATCAAGGCTTCAAGGATTTTACCTTTCCATA 720
DB 796 CTGCGCTTCGCGAGGCTTCGAGGAGATCAAGGCTTCAAGGATTTTACCTTTCCATA 855
QY 721 GCAGATCATATTAGAGTTCGGAATGCAAAATACAGTGTGAAGAACTCACCCCA 780
DB 856 GCAGATCACTATGTGAAGTTCGGAATGCAAAATACAGTGTGAAGAACTCACCCCA 915
```

```
QY 781 GTTATAGGAGGCTATCCGGTTGAGAAATTGTGGCTACATGTATCATTTAGTTCAGTTT 840
DB 916 GTCATAGGAGGCTATCCCGTGGAGAAATTGTGGGAGCAATGTACCATTTTACAGTTT 975
QY 841 GCTATTATTAAGTTTGAAGACCTGGAAGATGAGAGCCCTGTGCAAGTACAGTATCTGCTC 900
DB 976 GCGTATTACAAGTTGAATGATGTGAAGATGAGAGCCCTGTGCAAGTACAGTATCTGCTC 1035
QY 901 TTTGATCAGATGACAAAGTTCATGACAGAGAACCTGTGTTATTACAGTACACAGGAC 960
DB 1036 TTTGACAGAGTACAGAGGTTATGCAACAGAACCTGTGTTACTATCAGTACACAGGAC 1095
QY 961 ACTTGGGCGCTTCTCGATGAGCACTTCCAGCCAGCTGAGAGTTCAGTTCCTTTAAT 1020
DB 1096 AAGTGGGCGCTTCTCGATGAGCACTTCCAGCCAGCTGAGAGTTCAGTTCCTTTAAT 1155
QY 1021 GTGACCACTCCAGAGGAGCTGTATGACTTCTTCTAGGAGAAATATATGATGATGAT 1080
DB 1156 GTGACGAGCTCCAGAGAGAACTGTACGACTTGTGAGAGCGAGAGTTC 1215
QY 1081 GAGGAGAGGTTGTGGAATATGTGATGACCTTCTTGAACCTGGAGGAGAC 1130
DB 1216 GAGGAGAGGTTGTGAGTATGTGAGCACTTGTGAGAGCGAGAGTTC 1265
RESULT 5
BX425905 1201 bp mRNA linear EST 15-MAY-2003
BX425905 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
CS0DA004YM07 5-PRIME, mRNA sequence.
ACCESSION BX425905
VERSION BX425905.1 GI:30774508
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2233.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0AA004AG04QPI&cluster=2233.f. Contact :
Peng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0AA004AG04QPI.
Location/Qualifiers
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DA004YM07"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/note="vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with NotI and cloned into
the NotI and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 250 a 334 c 314 g 248 t 55 others
ORIGIN
```

```
Query Match 78.8%; Score 893.8; DB 13; Length 1201;
Best Local Similarity 93.4%; Pred. No. 2.6e-206;
Matches 944; Conservative 28; Mismatches 31; Indels 8; Gaps 4;
```



QY	8	CCCAATAGACGCTACAGCTTCCGAGCTTCCACGGGACGAGCTGATCGCGCTCGAGT	67
Db	112	CCCAATAGACGCTACAGCTTCCGAGCTTCCACGGGACGAGCTGATCGCGCTCGAGT	171
QY	68	CGGCTACCGGACGCGCTGGACAAATGACAGGGGACGACTGGGCGGAGAGCTGCGCT	127
Db	172	CGGCTACCGGACGCGCTGGACAAATGACAGGGGACGACTGGGCGGAGAGCTGCGCT	231
QY	128	ACCTGAGATACGCTCGGCTGACCGCTGCTGCGGACGAGGAGCTTCTGCCACC	187
Db	232	ACCTGAGATACGCTCGGCTGACCGCTGCTGCGGACGAGGAGCTTCTGCCACC	289
QY	188	GCAACTGACGCGCGCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	247
Db	290	GCAACTGACGCGCGCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	349
QY	248	TGCGGCTCTTTCGGGGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	307
Db	350	TGCGGCTCTTTCGGGGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	409
QY	308	TGCCAGCTTCCGCGAGTCCGAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	367
Db	410	TGCCAGCTTCCGCGAGTCCGAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	469
QY	368	AGCCTACAGTTCCTGAGTTCGTTACTTCAAGGCAAAATAATCTCCCAAGCCATCG	427
Db	470	AGCYTACAGTTCCTGAGTTCGTTACTTCAAGGCAAAATAATCTCCCAAGCCATCG	529
QY	428	CGCTGCTCACACCTTCTACTGAAGCATCTGATGAGCAATGATGAAGGAAACATGG	487
Db	530	CGCTGCTCACACCTTCTACTGAAGCATCTGATGAGCAATGATGAAGGAAACATGG	589
QY	488	CATATTAAGAGCTGCTGCTGCGGAGCTACATTAAGAGCTGCTGCTGCTGCTGCTGCT	547
Db	590	CATATTAAGAGCTGCTGCTGCGGAGCTACATTAAGAGCTGCTGCTGCTGCTGCTGCT	649
QY	548	ATGAAGCTGTTTATCCGAGCTGCGGCGGATACACGGTGAGAACTGGAGACATCCA	607
Db	650	ATGAAGCTGTTTATCCGAGCTGCGGCGGATACACGGTGAGAACTGGAGACATCCA	709
QY	608	TCACAGATGAGTGGCCCTTCCGACTTCTTCAAAGCCTTTTACGAGTGTCTGCGAG	667
Db	710	TCACAGATGAGTGGCCCTTCCGACTTCTTCAAAGCCTTTTACGAGTGTCTGCGAG	769
QY	668	CCTGCGAGGTTCCAGGAGATCAAGGACTTCAAGGATTTTACCTTTCCATAGCATC	727
Db	770	CCTGCGAGGTTCCAGGAGATCAAGGACTTCAAGGATTTTACCTTTCCATAGCATC	829
QY	728	ATTATGTAGAAGTTCCTGGAATGCAAAATACAGTGTGAAGAGAACTCACCCGCTTATAG	787
Db	830	ATTATGTAGAAGTTCCTGGAATGCAAAATACAGTGTGAAGAGAACTCACCCGCTTATAG	889
QY	788	GAGGTATCCGGTTGAGAAATTTGTGGCTACATGTATCATTTACGAGTTTCCCTATT	847
Db	890	GAGGTATCCGGTTGAGAAATTTGTGGCTACATGTATCATTTACGAGTTTCCCTATT	949
QY	848	ATPAGTTGAAAGACCTGAAGATGACAGCCCTGTCAGTCAAGTATCTGCTTTTGTATC	907
Db	950	ATPAGTTGAAAGACCTGAAGATGACAGCCCTGTCAGTCAAGTATCTGCTTTTGTATC	1008
QY	908	AGAATGACAGGTATGACAGAGAACCTTGTGTATTACCATACACAGGAGACCTTTGGG	967
Db	1009	AGAATGACAGGTATGACAGAGAACCTTGTGTATTACCATACACAGGAGACCTTTGGG	1063
QY	968	GCTCTCGAGTACGACTTCCAGCCAGACCTGAGGAGCTTCACTTTTAA 1018	
Db	1064	GGSYTTGGTAKVCTYTCACCCARACTTAACARTYAMWYTTTWTWTW 1114	

BASE COUNT

253 a 363 c 315 g 216 t

1 others

ORIGIN

Query Match 78.4%; Score 889; DB 13; Length 1148;  
Best Local Similarity 95.9%; Pred. No. 3.7e-205;  
Matches 943; Conservative 2; Mismatches 32; Indels 6; Gaps 3;

QY

1

GGCGCGCCCAATACGAGCTACAGCTTCCGAGCTTCCGAGCTTCCGAGCTGATCGCG

60

Db

66

GGCGCGCCCAATACGAGCTTCCGAGCTTCCGAGCTTCCGAGCTGATCGCG

125

QY

61

CTCGAGTCCGCTACCGGACGCGCTGGACAACTACAGCGGAGCAGCTGGGCGGAGC

120

Db

126

CTCGAGTCCGCTACCGGACGCGCTGGACAACTACAGCGGAGCAGCTGGGCGGAGC

185

QY

121

STKGGTACTCGAGATCAGCTGCGGCTGACCGCTTCTGCGGAGCAGGAGGCTTC

180

Db

186

GTGGGTACTCGAGATCAGCTGCGGCTGACCGCTTCTGCGGAGCAGGAGGCTTC

245

QY

181

TGCACCGCAACTGACGAGCGCGCGGAGCAGCGCGGAGCAGCGCGGCTTCCGAGCTAT

240

Db

246

TGCACCGCAACTGACGAGCGCGCGGAGCAGCGCGGAGCAGCGCGGCTTCCGAGCTAT

305

QY

241

CCGAGCTCGGCTCTTTCGGGGCTGCTGCGCGCGGAGCTGCTCAAGGCTGCGAG

300

Db

306

CCGAGCTCGGCTCTTTCGGGGCTGCTGCGCGCGGAGCTGCTCAAGGCTGCGAG

365

QY

301

CAGGGCTCCGAGCTTCCGCGAGTCCGAGCGGAGGCTGCTGGGAGCTTCAG

360

Db

366

CAGGGCTCCGAGCTTCCGCGAGTCCGAGCGGAGGCTGCTGGGAGCTTCAG

425

QY

361

CGCGCGAGCGCTCAAGTTCCTGAGTTCGTTACTTCAAGGCAAAATAATCTCCCAA

420

Db

426

CGCGCGAGCGCTCAAGTTCCTGAGTTCGTTACTTCAAGGCAAAATAATCTCCCAA

485

QY

421

GCCATCGCGCTGCTCACACCTTCTTACTGAGCATCTGATGAGCAATGATGAGG

480

DEFINITION AGENCOURT 10127451 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:6502891

5', mRNA Sequence.

ACCESSION BU902192

VERSION BU902192.1 GI:24084105

KEYWORDS EST.

ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 1148)

AUTHORS NIH-MGC htbp://mgi.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLAM14059 row: k column: 20

High quality sequence stop: 754.

Location/Qualifiers

1..1148

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:6502891"

/tissue\_type="leiomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_71"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site.1: NotI; Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb."

BASE COUNT 253 a 363 c 315 g 216 t

ORIGIN

Query Match 78.4%; Score 889; DB 13; Length 1148;  
Best Local Similarity 95.9%; Pred. No. 3.7e-205;  
Matches 943; Conservative 2; Mismatches 32; Indels 6; Gaps 3;

QY 1 GGCGCGCCCAATACGAGCTACAGCTTCCGAGCTTCCGAGCTGATCGCG 60

Db 66 GGCGCGCCCAATACGAGCTTCCGAGCTTCCGAGCTGATCGCG 125

QY 61 CTCGAGTCCGCTACCGGACGCGCTGGACAACTACAGCGGAGCAGCTGGGCGGAGC 120

Db 126 CTCGAGTCCGCTACCGGACGCGCTGGACAACTACAGCGGAGCAGCTGGGCGGAGC 185

QY 121 STKGGTACTCGAGATCAGCTGCGGCTGACCGCTTCTGCGGAGCAGGAGGCTTC 180

Db 186 GTGGGTACTCGAGATCAGCTGCGGCTGACCGCTTCTGCGGAGCAGGAGGCTTC 245

QY 181 TGCACCGCAACTGACGAGCGCGCGGAGCAGCGCGGAGCAGCGCGGCTTCCGAGCTAT 240

Db 246 TGCACCGCAACTGACGAGCGCGCGGAGCAGCGCGGAGCAGCGCGGCTTCCGAGCTAT 305

QY 241 CCGAGCTCGGCTCTTTCGGGGCTGCTGCGCGCGGAGCTGCTCAAGGCTGCGAG 300

Db 306 CCGAGCTCGGCTCTTTCGGGGCTGCTGCGCGCGGAGCTGCTCAAGGCTGCGAG 365

QY 301 CAGGGCTCCGAGCTTCCGCGAGTCCGAGCGGAGGCTGCTGGGAGCTTCAG 360

Db 366 CAGGGCTCCGAGCTTCCGCGAGTCCGAGCGGAGGCTGCTGGGAGCTTCAG 425

QY 361 CGCGCGAGCGCTCAAGTTCCTGAGTTCGTTACTTCAAGGCAAAATAATCTCCCAA 420

Db 426 CGCGCGAGCGCTCAAGTTCCTGAGTTCGTTACTTCAAGGCAAAATAATCTCCCAA 485

QY 421 GCCATCGCGCTGCTCACACCTTCTTACTGAGCATCTGATGAGCAATGATGAGG 480



FEATURES		Location/Qualifiers	
source		1..1652	
		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="C57BL/6J"	
		/db_xref="PANTOM DB:5730529N23"	
		/db_xref="MGI:1898560"	
		/db_xref="taxon:10090"	
		/clone="5730529N23"	
		/issue_type="whole body"	
		/clone_lib="RIKEN full-length enriched mouse cDNA library"	
		/dev_stage="8 days embryo"	
CDS		77..1279	
		/note="unnamed protein product; cartilage associated protein (MGI:1891221, GB NM_019922, evidence: BLASTN, 99%, match=1655)"	
		putative"	
		/codon_start=1	
		/protein_id="BAB30938.1"	
		/db_xref="GI:12857227"	
		/db_xref="MGI:1891221"	
		/translation="MGPSPTAALLVLCVCAPTGPGGOVERYSFSPFDELMPLE SAYRHALQYGEHWAEVGVLEVSRLRLHLLRBRQAFCHENCSSAAFPAPAPASH AELRFGVLRARCKKQSLPAFROSQSRSLADFOQREFYKLFQYAFKRNGL PRALAAHTYLLKHDDEMRKNMEYYKSLPGEADHIKDLSTKYSLSLFVRAVFNGL ENWRTSIDMELALPDFKAFYECIAACEGVAETKPKDFYLSIADHYFVELECKIRC BETLTPVIGGVPEVFATVHYLQFAYYKLNIAAPCAVSYLLDFQSDRVNQNL VYQYHRDKWGLSDHFPQPEAVQFENVVTLQKELYDFAQEHLMDDDEGEVVEYVD LLETESA"	
BASE COUNT	338 a	503 c	445 g 366 t
ORIGIN			
	Query Match	78.1%;	Score 886; DB 11; Length 1652;
	Best Local Similarity	86.4%;	Pred. No. 2.4e-204;
	Matches	976;	Conservative 2; Mismatches 152; Indels 0; Gaps 0;
QY	1	GGGCGGCCCATATGACAGCTTACAGCTTCGACGCTTCCACGGAGAGCTGATGCGG	60
DB	143	GGGCGGGCAGTAGAGCGCTACAGCTTCGACGCTTCCGCGGGAGAGCTGATGCGG	202
QY	61	CTCGAGTCGGCTACCGGACCGCGCTGGACAGTACAGCGCGGAGCACTGGCGCGAGAGC	120
DB	203	CTCGAGTCGGCTACCGGACCGCGCTGCACAGTACAGCGCGGAGCACTGGCGCGAGAGC	262
QY	121	STKGCTACTCGAGATCAGCTTGGCGGTGCAACCGCTTGTGCGCGAGCAGCGAGCCTTC	180
DB	263	GTGGGTACTCGAGTGAGCTTGGCGGTGCAACCGCTTGTGCGCGAGCAGCGAGCCTTC	322
QY	181	TGCCACCGCACTGACGCGCGCGCGGACCGCGGAGCGCGCGGCTCGCCAGCTAT	240
DB	323	TGCCACCGCACTGACGCGCGCGGACCGCGGAGCGCGGCTCGCCAGCTAT	382
QY	241	CCGAGCTCGGCTTCTGGGGGCTGTGTCGCGCGGCGCACTGCTCAAGCGCTGCAAG	300
DB	383	GCGCAACTACGCGCTTCTGGCAGCGTGTCTGCGCGCGGCGAGTCTCAAGCGCTGCAAG	442
QY	301	CAGGCGCTCGAGCTTCCGCGAGTCCGCGAGCGCGGAGCGGCTGCGCGAGCTTCCAG	360
DB	443	CAGGCGCTCGGCGCTTCCGCGAGTCCGCGAGCGCGGAGCGGCTGCGCGAGCTTCCAG	502
QY	361	CGCGCGAGCGCTTCAAGTTCCTGCGAGTTCGCTTACTTCAAGCGCAAAATCTCCCGAAA	420
DB	503	CAGCGGAGCGCTTCAAGTTCCTGCGAGTTCGCTTACTTCAAGCGCAAAATCTCCCGAAA	562
QY	421	GCATCGCGCTGCTCACACCTTTCTACTGAAGCATCTGTATGAGCAATATGATGAGAGG	480
DB	563	GCCATCGCTGCGGCTCACACCTTTCTTCTGAAGCATCTGTATGAGCAATATGATGAGAGG	622
QY	481	AACATGGCATTTATTAAGAGCGCTGCTGCTGCGAGGACTACATTAAGAGCGCTGGAACC	540
DB	623	AACATGGAGTATTTAAGAGCTTGTCTGAGCGGAGGACCATTTAAGAGCTTGAAGACC	682
QY	541	AAGTCATATGAAAGCGCTGTTCTATCGAGCAGTTCGCGGGCATCAACGGGTGAGAACTGGAGA	600
Db	583	AGTCTGACGAGAGCTGTTTCTCGTGGGTGGGGCTACACGGGAGAGCTGGAGA	742
QY	601	ACATCCATCACAGACATGAGAGCTGGCCCTTCCGAGCTTCTTCAAGAGCTTTTACAGTGT	660
Db	743	ACGTCCATTTCCGACATGAGAGCTCGCGCTTCCGAGCTTCTTCAAGAGCTTTTACAGTGT	802
QY	661	CTCGAGCGCTGGAGGGTTCCAGGAGATCAAGGAGTTCACAGGATTTCTACCTTTCCATA	720
Db	803	CTGGCTGCTCGAGGGGGTCCGGAGATCAAGGATTTCTACCTTTCCATA	862
QY	721	GCAGATCATTTATGAGAGTTCGGAATGCAAAATACAGTGTGAAGAGAACTTCAACCCCA	780
Db	863	GCAGATCATTTATGAGAGTTCGGAATGCAAAATACAGTGTGAAGAGAACTTCAACCCCA	922
QY	781	GTATAGGAGGCTATCCGGTTGAGAAATTTGGGTACCATGATATCATTTACCTTTCAGTTT	840
Db	923	GTATAGGAGGCTATCCGGTTGAGAAATTTGGGTACCATGATATCATTTACCTTTCAGTTT	982
QY	841	GCCTATTATAAGTTGAAGACCTGGAAGAAATGAGAGGCTTCCAGCGGCTTTCAGTCTATCTGTC	900
Db	983	GCCTATTATAAGTTGAAGACCTGGAAGAAATGAGAGGCTTCCAGCGGCTTTCAGTCTATCTGTC	1042
QY	901	TTTGATCAGAAATGACAGAGCTCATGAGCAGAGAGACCTGGTGTATTACAGTACACAGGGAC	960
Db	1043	TTTGATCAGAGTACAGAGGCTCATGAGCAGAGAGACCTGGTGTATTACAGTACACAGGGAC	1102
QY	961	ACTTGGGGCTCTTCGGATGAGCACTTCCAGCGGACAGACCTTGAAGCAGTTCAGTCTTTTAT	1020
Db	1103	AAGTGGGGCTCTTCGGATGAGCACTTCCAGCGGACAGACCTTGAAGCAGTTCAGTCTTTTAT	1162
QY	1021	GTGACACACTCCAGAGAGCTGTATGACTTGTGTAAGAAAATATATGATGATGAT	1080
Db	1163	GTGACACACTCCAGAGAGCTGTATGACTTGTGTAAGAAAATATATGATGATGAT	1222
QY	1081	GAGGAGAGAGTTTGTGAATATGAGATGACCTTCTTGAAGCTGGAGAGAGC	1130
Db	1223	GAGGAGAGAGTTTGTGAATATGAGATGACCTTCTTGAAGCTGGAGAGAGC	1272
RESULT 8			
LOCUS	AL546910	1075 bp	mRNA linear EST 31-MAY-2003
DEFINITION	AL546910 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0D1032YH24 5-PRIME, mRNA sequence.		
ACCESSION	AL546910		
VERSION	AL546910.2	GI:31268743	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1. (bases 1 to 1075)		
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished		
COMMENT	On Feb 15, 2001 this sequence version replaced gi:12880487.		
	Contact: Genoscope - Centre National de Sequenage		
	Genoscope - Centre National de Sequenage		
	BP 191 91006 EVRY cedex - France		
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	Library was constructed by Life Technologies, a division of		
	Invitrogen. This sequence belongs to sequence cluster 2233.f For		
	more information about this cluster, see		
	http://www.genoscope.cns.fr/		
	csi-bin/cluster.cgi?seq=CS0D1032YH24&cluster=2233.f. Contact :		
	Feng Liang Email : fliang@lifetech.com URL :		
	http://fulllength.invitrogen.com/Invitrogen Corporation 1600		
	Paradise Avenue Genoscope sequence ID : CS0D1032YH24Q1.		
FEATURES			
source			
	1..1075		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		



```
Db 372 GCGCCTCTTTCGGGGGCTGTGCGCGCGCGCACTGCCTCAAGCGTGTGAAGCAGGGCT 431
Qy 309 GCAGCCTTCCGCGAGTCCAGAGCCAGCGAGGCTGTGGCGAGCTTCCAGCGCGCGA 368
Db 432 GCAGCCTTCCGCGAGTCCAGAGCCAGCGAGGCTGTGGCGAGCTTCCAGCGCGCGA 491
Qy 369 GCCTCAAGTCTCTGAGTTCCTTACTTCAAGGCAAAATATCTCCCAAGCCATGCG 428
Db 492 GCCTCAAGTCTCTGAGTTCCTTACTTCAAGGCAAAATATCTCCCAAGCCATGCG 551
Qy 429 CGTGTCTCACACCTTCTACTGAGCATCTGATGAGCAATGATCAAGAGCAATGCG 488
Db 552 CGTGTCTCACACCTTCTACTGAGCATCTGATGAGCAATGATCAAGAGCAATGCG 611
Qy 489 ATATTATAAGAGCTGCTGCTGCGGAGGACTACATTAAGAGCTTGGAAACCAAGTCATA 548
Db 612 ATATTATAAGAGCTGCTGCTGCGGAGGACTACATTAAGAGCTTGGAAACCAAGTCATA 671
Qy 549 TGAAGCCTTTCATCGGAGCAGTGGGGGATACAAAGGTGAGAACTGGAGAACTCCAT 608
Db 672 TGAAGCCTTTCATCGGAGCAGTGGGGGATACAAAGGTGAGAACTGGAGAACTCCAT 731
Qy 609 CACAGACATGAGCTGGCCCTTCCGAGCTTCTTCAAGGCTTTCAGAGTGTCTGCGAG 668
Db 732 CACAGACATGAGCTGGCCCTTCCGAGCTTCTTCAAGGCTTTCAGAGTGTCTGCGAG 791
Qy 669 CTCGAGGGTTCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGAGATCA 728
Db 792 CTCGAGGGTTCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGAGATCA 851
Qy 729 TTATGTAGAGTTCGGAATGCAAAATACAGTGTGAAGAGAACCTCACCCAGTTATAGG 788
Db 852 TTATGTAGAGTTCGGAATGCAAAATACAGTGTGAAGAGAACCTCACCCAGTTATAGG 911
Qy 789 AGCTATCCGGTTGAGAAATTTG-TGGTACCATGTATCATTTACCTTTCAGTTTGGCTATT 847
Db 912 AGCTATCCGGTTGAGAAATTTG-TGGTACCATGTATCATTTACCTTTCAGTTTGGCTATT 971
Qy 848 ATAAGTGTACGACCTGAGAGATGAGCCCTTGGTGCAGTACGATCTGCTCTTTGATC 907
Db 972 ATAAGTGTACGACCTGAGAGATGAGCCCTTGGTGCAGTACGATCTGCTCTTTGATC 1031
Qy 908 AGAATGACCAAG 918
Db 1032 AGAWKEAAGG 1042
```

```
RESULT 10
AL541167
LOCUS AL541167 Homo sapiens PLACENTA 1201 bp mRNA linear EST 12-MAY-2003
DEFINITION S-PRIME, mRNA sequence.
ACCESSION AL541167
VERSION AL541167.2 GI:30545079
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12871971.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 9106 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2233.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
```

```
cgi-bin/cluster.cgi?seq=CS0DE005DC06QPI&cluster=2233.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DE005DC06QPI.
FEATURES
source
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE005YF12"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dt) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 258 a 335 c 313 g 237 t 58 others
ORIGIN
Query Match 76.3%; Score 865.6; DB 9; Length 1201;
Best Local Similarity 98.1%; Pred. No. 1.9e-199;
Matches 895; Conservative 10; Mismatches 4; Indels 3; Gaps 3;
Qy 8 CCCAATACGACGCTACAGCTTCCGAGCTTCCACGGGACGAGCTGATCCGCTCGAGT 67
Db 139 CCCAATACGACGCTACAGCTTCCGAGCTTCCACGGGACGAGCTGATCCGCTCGAGT 198
Qy 68 CGGCTTACCAGCGCTGAGCAAGTACAGCGCGAGCACTGGCGGAGAGCTGAGCT 127
Db 199 CGGCTTACCAGCGCTGAGCAAGTACAGCGCGAGCACTGGCGGAGAGCTGAGCT 258
Qy 128 ACCTGGAGATCAGCTCGGCTGCAAGTACAGCGCGAGCACTGGCGGAGAGCTGAGCT 187
Db 259 ACCTGGAGATCAGCTCGGCTGCAAGTACAGCGCGAGCACTGGCGGAGAGCTGAGCT 317
Qy 188 GCAACTCAGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 247
Db 318 GCAACTCAGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 377
Qy 248 TGGCGCTTTCGGGGGCTTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 307
Db 378 TGGCGCTTTCGGGGGCTTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 437
Qy 308 TGGCGAGCTTCCGCGAGTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 367
Db 438 TGGCGAGCTTCCCGAGTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 497
Qy 368 AGCCTTACAAGTTCCTGAGTTCGCTTACTTCAAGCAAAATATCTCCCAAGCCATCG 427
Db 498 AGCCTTACAAGTTCCTGAGTTCGCTTACTTCAAGCAAAATATCTCCCAAGCCATCG 557
Qy 428 CGCTGTCTACACCTTCTACTGAGCATCTGATCAGCAATGATGAGAGAGAACTCG 487
Db 558 CGCTGTCTACACCTTCTACTGAGCATCTGATGAGCAAAATGATGAGAGAGAACTCG 617
Qy 488 CATATTATAAGAGCTGCTGCTGCGAGGACTACATTAAGACCTGGGAAACCAAGTCAT 547
Db 618 CATATTATAAGAGCTGCTGCTGCGAGGACTACATTAAGACCTGGGAAACCAAGTCAT 677
Qy 548 ATGAAGCCTTTCATCCGAGCAGTGGGGGATACAAACGCGTGAAGAACTGGAGAACTCCA 607
Db 678 ATGAAGCCTTTCATCCGAGCAGTGGGGGATACAAACGCGTGAAGAACTGGAGAACTCCA 737
Qy 608 TCACAGACATGGAGCTGGCCCTTCCGAGCTTCTTCAAGCCTTTTACGAGTGTCTCGCAG 667
Db 738 TCACAGACATGGAGCTGGCCCTTCCGAGCTTCTTCAAGCCTTTTACGAGTGTCTCGCAG 797
Qy 668 CTTGCGAGGTTCCAGGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAGATC 727
Db 798 CTTGCGAGGTTCCAGGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAGATC 857
Qy 728 ATTATGTAGAGTTCCTGAGTGCAGAAATACAGTGTGAAGAGAACTCACCCCGATTATAG 787
```

```

Db      858 ATTATGTAGAAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAACCTCACCCAGTTATAG 917
QY      788 GAGGCTATCCGGTTGAGAAATTTGGGTACCATATATCATTTGAGTTTGCCTATT 847
Db      918 GAGGCTATCCGGTTGAGAAATTTGGGTACCATATATCATTTGAGTTTGCCTATT 977
QY      848 ATAAGTTGAACGACCTGAAGAAATGAGCCCTGTGAGTACGATATCTCTTTGATC 907
Db      978 ATAATTTG-AMGACCTKGAAATGAG-CCCTGTGAGTACGATATTTGCTTTGWTM 1035
QY      908 AGAATGACAAAG 919
Db      1036 AGAATGADAAG 1047

RESULT 11
LOCUS   AL554750
DEFINITION AL554750 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
ACCESSION AL554750
VERSION   AL554750.2 GI:31276560
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12895833.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2233.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1086DE02P1&cluster=2233.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D1086DE02P1.

FEATURES
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1086YJ04"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 272 a 329 c 309 g 250 t 41 others
ORIGIN

Query Match 71.6%; Score 811.6; DB 9; Length 1201;
Best Local Similarity 95.1%; Pred. No. 2.5e-186;
Matches 871; Conservative 18; Mismatches 18; Indels 9; Gaps 5;

QY      8 CCCAATAGAACCTACAGC-TTCCGCGAGCTTCCACGGGACGAGCTGATCGCGTCCAG 66
Db      151 CCCAATAGAACCTACAGCTTCCGAGCTTCCACGGGACGAGCTGATCGCGTCCAG 210
QY      67 TCGGCTACCGGACGCGGTGACAAAGTACAGCGCGGACACTGG--GCCGAGAGCTG 124
Db      211 TCGGCTACCGGACGCGGTGACAAAGTACAGCGCGGACACTGGCGCGGAGAGCTGG 270
QY      125 GCTACTGTGAGATCAGCTTGGGCTGCACCGCTTCTGCTGGCGACAGCGAGCCTTCTGCC 184

```

```

Db      271 GCTACTGTGAGATCAGCCTTGGGCTGCACCGTTGCTGGCGACACTAKTCTTCTTGGC 330
QY      185 ACCGCAACTGCAGCGCGCGCGCGAGCCC--GCCGCCGGCTCGCCAGCTATCC 242
Db      331 ACCGCAACTTCAKCTCCGCGCGCGCAKCCKXCCCTGTCCGCCGCTCGCCAGCTATCC 390
QY      243 CGAGCTGGCGCTTCTTGGGGGCTTGTGCGCGCGGCGACTGCTCAAGCGCTGCAAGCA 302
Db      391 CGAGCTGGCGCTTCTTGGGGGCTTGTGCGCGCGCGCACTGCTCAAGCGCTGCAAGCA 450
QY      303 GGGCTTCCAGAGCCTTCCGCCAGTCCCGAGCCCGAGCGGCTGCTGGCGGACTTCCAGCG 362
Db      451 GGGCTTCCAGAGCCTTCCGCCAGTCCCGAGCAGTCCAGCAGTCCCGGAGCTTCCAGCG 510
QY      363 CCGCGAGCCTTACAGTTCCTGAGTTCCTTCAAGGCAAAATAATCTTCCCAAGC 422
Db      511 CCGCGAGCCTTACAGTTCCTGAGTTCCTTCAAGGCAAAATAATCTTCCCAAGC 570
QY      423 CATCGCGCTGCTCACACCTTTCTACTGAAGCATCTGATGAGAAATCATGAAGAGAA 482
Db      571 CATCGCGCTGCTCACACCTTTCTACTGAAGCATCTGATGAGAAATCATGAAGAGAA 630
QY      483 CATGGCATATTATAGAGCCTGCTGCTGCGGAGCTACATTAAAGACCTCGAAACCAA 542
Db      631 CATGGCATATTATAGAGCCTGCTGCTGCGGAGCTACATTAAAGACCTCGAAACCAA 690
QY      543 GTCATATGAAGCCTGTTTCATCGAGCAGTGGGGCATACAAAGTGAAGACTGGAGAAC 602
Db      691 GTCATATGAAGCCTGTTTCATCGAGCAGTGGGGCATACAAAGTGAAGACTGGAGAAC 750
QY      603 ATCCATCAGACATGAGCTGCGCTTCCCGACTTCTTCAAAGCCTTTTACGAGTGTCT 662
Db      751 ATCCATCAGACATGAGCTGCGCTTCCCGACTTCTTCAAAGCCTTTTACGAGTGTCT 810
QY      663 CGAGCCTGCGAGGTTTCCAGGAGATCAAGACTTCAAGGATTTCTACTTCCATAGC 722
Db      811 CGAGCCTGCGAGGTTTCCAGGAGATCAAGACTTCAAGGATTTCTACTTCCATAGC 870
QY      723 AGATCATTATGTAAGTCTGGAATGCAAAATACAGTGTGAAGAGAACTCACCCAGT 782
Db      871 AGATCATTATGTAAGTCTGGAATGCAAAATACAGTGTGAAGAGAACTCACCCAGT 930
QY      783 TATAGAGGCTATCCGGTTGAGAAATTTGTGCTACCATGTATCATATTGCGAGTTGC 842
Db      931 TATAGAGGCTATCCGGTTGAGAAATTTGTGCTACCATGTATCATATTGCGAGTTGC 990
QY      843 CTATTATAGTTGAACGACCTGAGAAATGCAAGCCTTGGAGCCCTTGGAGTATCTGCTT 902
Db      991 CTATTATAGTTGAACGAC--TGAARATGACCCCTTGGAGTMA-STATCTGCTCTB 1046
QY      903 TGATCAGATGACAAG 918
Db      1047 TGATCAGATGACAG 1062

RESULT 12
LOCUS   AL543090
DEFINITION AL543090 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
ACCESSION AL543090
VERSION   AL543090.2 GI:31264937
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12875568.

```



Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 2233.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D1002CH09QPl&cluster=2233.f. Contact:  
Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID: CS0D1002CH09QPl.  
Location/Qualifiers

FEATURES  
source

1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1002YP17"  
/tissue\_type="PLACENTA"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 272 a 333 c 326 g 237 t 33 others

Query Match 71.2%; Score 807.8; DB 9; Length 1201;  
Best Local Similarity 98.8%; Pred. No. 2.1e-185;  
Matches 839; Conservative 5; Mismatches 2; Indels 3; Gaps 3;

QY 8 CCCAATACGAACGCTACAGCTTCGGAGCTTCCACGGGACGAGCTGATCGCGCTCGAGT 67  
DB 138 CCCAATACGAACGCTACAGCTTCGGAGCTTCCACGGGACGAGCTGATCGCGCTCGAGT 197  
QY 68 CGGCTACCGGACGCGCTCGAAGTACAGCGGCGAGCACTGGGCGGAGAGCTGCGT 127  
DB 198 CGGCTACCGGACGCGCTCGAAGTACAGCGGCGAGCACTGGGCGGAGAGCTGCGT 257  
QY 128 ACCTGGAGATACGCTCGGCTGACCGCTTCTGGCGGACGAGGCTTCTGCACC 187  
DB 258 ACCTGGAGATACGCTCGGCTGACCGCTTCTGGCGGACGAGGCTTCTGCACC 317  
QY 188 GCAACTGACGCGCGCGCGAGCGCGCGCGCGCGCGCTCGCAGCTATCCCGAGC 247  
DB 318 GCAACTGACGCGCGCGCGAGCGCGCGCGCGCGCGCTCGCAGCTATCCCGAGC 377  
QY 248 TGGGCTTCTGGGCGGCTGCTGGGCGCGCGCGCGCGCTCAAGCGCTGCAAGCGGCG 307  
DB 378 TGGGCTTCTGGGCGGCTGCTGGGCGCGCGCGCGCGCTCAAGCGCTGCAAGCGGCG 437  
QY 308 TGGGCGCTTCCGCGAGTCCCGAGCGCGCGCGCGCGCTTGGCGGACTTCCAGCGCGCG 367  
DB 438 TGGGCGCTTCCGCGAGTCCCGAGCGCGCGCGCGCGCTTGGCGGACTTCCAGCGCGCG 497  
QY 368 AGCCCTACAGTTCCTCGAGTTCCTTACCTTCAAGGCAATATCTCCCAAGCCATCG 427  
DB 498 AGCCCTACAGTTCCTCGAGTTCCTTACCTTCAAGGCAATATCTCCCAAGCCATCG 557  
QY 428 CCGCTGCTCACACCTTCTACTGAAGCATCTGTATGACGAAATGATGAAGAGGAACATGG 487  
DB 558 CCGCTGCTCACACCTTCTACTGAAGCATCTGTATGATGAAGAGGAACATGG 617  
QY 488 CATATTATAGAGCTCGCTGGTGGCGGAGCTACATTAAGACCTGGAAACCAAGTCAT 547  
DB 618 CATATTATAGAGCTCGCTGGTGGCGGAGCTACATTAAGACCTGGAAACCAAGTCAT 677  
QY 548 ATGAAGCCCTTTCATCCGAGCGAGTGGGCGCATACACGCTGAGACTGGAGAACATCCA 607  
DB 678 ATGAAGCCCTTTCATCCGAGCGAGTGGGCGCATACACGCTGAGACTGGAGAACATCCA 737  
QY 608 TCACAGACATGGAGCTGGCGCTTCCCGACTTCTTCAAAGCTTCTTACGAGTGTCTCGCAG 667

Db 738 TCACAGACATGGAGCTGGCGCTTCCCGACTTCTTCAAAGCTTCTTACGAGTGTCTCGCAG 797  
QY 668 CCTCGAGGGTTCAGGGAGATCAAGGACTTCAAGGATTTCTACCTTCCATAGCAGATC 727  
Db 798 CCTCGAGGGTTCAGGGAGATCAAGGACTTCAAGGATTTCTACCTTCCATAGCAGATC 857  
QY 728 ATTATGATAGAGTTCGGAATGCAAAATACAGTGTGAAGAGAACTTCAACCCAGTTATAG 787  
Db 858 ATTATGATAGAGTTCGGAATGCAAAATACAGTGTGAAGAGAACTTCAACCCAGTTATAG 917  
QY 788 GAGGCTATCCGTTGAGAAATTTGTGGCTACCATGTATCATTTACTTGCAGTTTGCCTATT 847  
Db 918 G-ROCTATCCGTTGAG-ANTTTGTGGCTACCATGTATCATTTACTTGCAGTTTG-CTATT 974  
QY 848 ATAAAGTTGA 856  
Db 975 ATAAAGTTGA 983

RESULT 13  
BU838140

LOCUS  
DEFINITION AGNCOURT 8071749 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6088409  
5', mRNA sequence.

ACCESSION BU838140  
VERSION BU838140.1 GI:24022535

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (Bases 1 to 846)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: DCTD/BTP

cDNA Library Preparation: Rubin Laboratory

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LICM2326 row: e column: 14

High quality sequence stop: 664.

FEATURES

source

Location/Qualifiers

1..846  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6088309"  
/tissue\_type="melanotic melanoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_112"  
/note="Organ: Skin; Vector: pOTB7; Site: 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Library constructed by Ling Hong in the  
Laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

BASE COUNT 199 a 255 c 215 g 177 t

ORIGIN

Query Match 70.6%; Score 801; DB 13; Length 846;

Best Local Similarity 100.0%; Pred. No. 7.8e-184;

Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 CTACTCGAGATACGCTGCGGTGACCGCTTCTCTCGCGGACGAGCGGCTTCTGCCA 185

Db 1 CTACTCGAGATACGCTGCGGTGACCGCTTCTCTCGCGGACGAGCGGCTTCTGCCA 60







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: January 30, 2004, 05:50:18 ; Search time 294.825 Seconds  
(without alignments)  
10382.969 Million cell updates/sec

Title: US-09-729-674-1\_COPY\_132\_1265  
Perfect score: 1134  
Sequence: 1 99cgcgcgcacacgaacg.....tgsaactgagagaccacg 1134

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_19Jun03:\*

```
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1133.2	99.9	3871	20	Human secreted pro
2	1133.2	99.9	3871	22	Human cDNA encodin
3	1133.2	99.9	3871	24	Human polynucleoti
4	1128.4	99.5	1961	22	Human cDNA sequenc
5	705	62.2	1586	21	Human colon cancer
6	587.6	51.8	818	22	Human cDNA clone (
7	455.4	40.2	2347	24	Oesophagus cancer
8	406.4	35.8	408	21	Human colon cancer

9	156.4	13.8	1309	24	Human polynucleoti
10	156.4	13.8	2152	24	Human polynucleoti
11	156.4	13.8	2211	24	DNA encoding novel
12	156.4	13.8	2242	24	Human polynucleoti
13	156.4	13.8	2524	22	Human cDNA sequenc
14	156.4	13.8	2563	22	Human cDNA encodin
15	156.4	13.8	2585	22	Human cDNA sequenc
16	156.4	13.8	2993	22	Human cDNA sequenc
17	154.8	13.7	2600	21	Human tumour suppr
18	154.8	13.7	2829	21	Human tumour suppr
19	152.6	13.5	2577	22	Human secreted pro
20	150.4	13.3	2615	22	Human secreted pro
21	147.4	13.0	2322	21	Mouse tumour suppr
22	147.4	13.0	2416	21	DNA encoding novel
23	130.8	11.5	2127	24	Mouse tumour suppr
c	126	11.1	618	21	DNA encoding novel
25	121.4	10.7	325	22	Human polynucleoti
26	116.2	10.2	1026	24	Oligonucleotide fo
27	116.2	10.2	1026	24	Oligonucleotide fo
c	115.8	10.2	1025	24	Oligonucleotide fo
29	115.8	10.2	1025	24	Oligonucleotide fo
30	109.2	9.6	2665	25	cDNA downregulated
31	99.2	8.7	652	22	Primer specific fo
32	99.2	8.7	738	22	Human cDNA clone (
33	99	8.7	703	22	Human cDNA clone (
34	98.6	8.7	774	22	Human cDNA clone (
35	98.6	8.7	800	22	Human cDNA clone (
36	98.6	8.7	2789	22	Human cDNA clone (
37	97.8	8.6	2753	24	Human novel polynu
38	97.8	8.6	2753	24	Human novel polynu
39	87	7.7	349	21	Lung cancer associ
40	81.6	7.2	1025	24	Oligonucleotide fo
c	81.6	7.2	1025	24	Oligonucleotide fo
42	81.6	7.2	1026	24	Oligonucleotide fo
c	81.6	7.2	1026	24	Oligonucleotide fo
44	73.2	6.5	550	25	Human adipocyte se
45	65.6	5.8	681	23	DNA encoding novel

#### ALIGNMENTS

#### RESULT 1

AA60801

ID AA60801 standard; DNA; 3871 bp.

XX AA60801;

AC AA60801;

XX 09-AUG-1999 (first entry)

XX Human secreted protein encoding DNA (clone bd306-7).

XX Secreted protein; kidney; lung; brain; blood; testis; bone marrow;  
XX nutritional activity; cytokine; cell proliferation; immune stimulation;  
XX hematopoiesis regulation; tissue growth; thrombolytic; gene therapy;  
XX anti-inflammatory; tumour invasion; ss.

OS Homo sapiens.

XX WO9926961-A1.

XX 03-JUN-1999.

XX 24-NOV-1998; 98WO-US25149.

XX 23-NOV-1998; 98US-019786.

XX 26-NOV-1997; 97US-0066804.

XX (GENY) GENETICS INST INC.

XX Agostino MJ, Clark HF, Collins-Racie LA, Evans C;

XX Fechtel K, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX Steininger RJ, Treacy M, Wong GG;

XX WPI: 1999-357809/30.  
DR P-PSDB; AAY17219.  
XX  
PT New polynucleotides encoding secreted proteins  
XX  
XX Claim 1; Page 110-111; 133pp; English.  
XX  
CC The invention relates to secreted proteins (AAY17219-228) encoded by  
CC polynucleotides obtained from human fetal kidney, adult lung, adult  
CC kidney, adult brain, adult blood, adult testes, and fetal brain and  
CC murine adult bone marrow cDNA libraries. The secreted protein nucleic  
CC acid sequences (X6801-811) correspond to clones bd306-7, g1283-6,  
CC fk17-3, k213-2x, na316-1, nf93-20, np164-1, pe204-1, ya1-1 and yb-1,  
CC (all clones are deposited as ATCC 98599); The PNs and proteins are  
CC predicted to have biological activities which would make them suitable  
CC for treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, cytokine and cell  
CC proliferation/differentiation activity, immune stimulating (e.g. as  
CC vaccines) or suppressing activity, hematopoiesis regulating activity,  
CC tissue growth activity, activin/ inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour  
CC invasion suppressor activity, and tumour inhibition activity. The PNs are  
CC also stated to be useful for gene therapy.  
XX  
SQ Sequence 3871 BP; 965 A; 913 C; 1033 G; 958 T; 2 other;

Query Match 99.9%; Score 1133.2; DB 20; Length 3871;  
Best Local Similarity 100.0%; Pred. No. 98-244;  
Matches 1134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCGCGCCCAATACGAACTGACAGCTTCGCGAGCTTCCACGGGACGAGCTGATGCCG 60  
DB 132 GGGCGCGCCCAATACGAACTGACAGCTTCGCGAGCTTCCACGGGACGAGCTGATGCCG 191

QY 61 CTCGAGTCGGCTACCGGCAACGCGCTGACAAAGTACAGCGCGGAGCACTGGCGGAGAGC 120  
DB 192 CTCGAGTCGGCTACCGGCAACGCGCTGACAAAGTACAGCGCGGAGCACTGGCGGAGAGC 251

QY 121 STKGCTACCTGGAGATCAGCTCGGCTGCGCTGACCGCTGCTGCGCGACGAGGCGCTTC 180  
DB 252 STKGCTACCTGGAGATCAGCTCGGCTGCGCTGACCGCTGCTGCGCGACGAGGCGCTTC 311

QY 181 TGCCACCGCAACTGCGCGCGCGCGCGAGCGCGCGCGCGCGCGCTCGCCAGCTAT 240  
DB 312 TGCCACCGCAACTGCGCGCGCGCGCGCGAGCGCGCGCGCGCGCTCGCCAGCTAT 371

QY 241 CCGGAGCTGCGCTTCGCGGCGCTGCTGCGCGCGCGGCACTGCTCAAGCGCTGCAAG 300  
DB 372 CCGGAGCTGCGCTTCGCGGCGCTGCTGCGCGCGCGGCACTGCTCAAGCGCTGCAAG 431

QY 301 CAGGCGCTGCGAGCTTCGCGGAGTCCCGGAGCGCGGAGCGGCTGGCGGAGCTCCAG 360  
DB 432 CAGGCGCTGCGAGCTTCGCGGAGTCCCGGAGCGCGGAGCGGCTGGCGGAGCTCCAG 491

QY 361 CGCGCGAGCGCTTCAAGTTCCTGAGTTCCTGAGTTCCTGAGGCAATATCTCCCGAAA 420  
DB 492 CGCGCGAGCGCTTCAAGTTCCTGAGTTCCTGAGTTCCTGAGGCAATATCTCCCGAAA 551

QY 421 GCCATCGCGCTGCTCACACCTTCTACTGAGGATCTGATGACGAAATGATGAAGAGG 480  
DB 552 GCCATCGCGCTGCTCACACCTTCTACTGAGGATCTGATGACGAAATGATGAAGAGG 611

QY 481 AACATGGCATATTATAAGAGCTGCTGGTGGCGGAGCTACATTAAGACCTGGAAACC 540  
DB 612 AACATGGCATATTATAAGAGCTGCTGGTGGCGGAGCTACATTAAGACCTGGAAACC 671

QY 541 AAGTCATATGAAGCTTCTTCATCCGAGCTGCGGCAACACCGTGAGACTGGAGA 600  
DB 672 AAGTCATATGAAGCTTCTTCATCCGAGCTGCGGCAACACCGTGAGACTGGAGA 731

QY 601 ACATCCATCACAGACATGGAGCTGGCCCTTCCCGACTTCTTCAAAGCCTTTTACAGGT 660  
DB 732 ACATCCATCACAGACATGGAGCTGGCCCTTCCCGACTTCTTCAAAGCCTTTTACAGGT 791

QY 661 CTCGAGCGCTGCGAGGTTCCAGGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATA 720  
DB 792 CTCGAGCGCTGCGAGGTTCCAGGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATA 851

QY 721 GCAGATCATTTATGTAGAGTTCTTGAATGCAAAATACAGTGTGAAGAGAACTCACCCCA 780  
DB 852 GCAGATCATTTATGTAGAGTTCTTGAATGCAAAATACAGTGTGAAGAGAACTCACCCCA 911

QY 781 GTTATAGAGGCTATCCCGTTGAGAAATTTGTGGCTACCATGTATCATTTTCAGTTT 840  
DB 912 GTTATAGAGGCTATCCCGTTGAGAAATTTGTGGCTACCATGTATCATTTTCAGTTT 971

QY 841 GCCTATTATAGTTGAACGACCTGAAGAAATCAGCCCCCTGTGAGTCAGTATCTGCTC 900  
DB 972 GCCTATTATAGTTGAACGACCTGAAGAAATCAGCCCCCTGTGAGTCAGTATCTGCTC 1031

QY 901 TTTGATCAGATGACAAAGTTCATGCGAGAGAACCTGGTGTATTACCATCCAGTACGAGGAC 960  
DB 1032 TTTGATCAGATGACAAAGTTCATGCGAGAGAACCTGGTGTATTACCATCCAGTACGAGGAC 1091

QY 961 ACTTGGGCGCTCTCGGATGAGCACTTCCAGGCCAGCACCTGAAGCAGTTTCCTTTAAT 1020  
DB 1092 ACTTGGGCGCTCTCGGATGAGCACTTCCAGGCCAGCACCTGAAGCAGTTTCCTTTAAT 1151

QY 1021 GTGACACACTCCAGAGGAGCTGTATGACTTTGTAAGGAAATATATATGATGATGAT 1080  
DB 1152 GTGACACACTCCAGAGGAGCTGTATGACTTTGTAAGGAAATATATATGATGATGAT 1211

QY 1081 GAGGAGAGAGTTGTGGAATATGTGGATGACCTCTTGGAACTGGAGGAGACACG 1134  
DB 1212 GAGGAGAGAGTTGTGGAATATGTGGATGACCTCTTGGAACTGGAGGAGACACG 1265

RESULT 2  
AAS59207  
ID AAS59207 standard; cDNA; 3871 BP.  
XX  
AC AAS59207;  
XX  
XX 16-JAN-2002 (first entry)  
XX  
DE Human cDNA encoding a secreted protein bd306\_7.  
XX  
XX Human; secreted protein; ss; antiinflammatory; immunosuppressive;  
XX neutrotic; neuroprotective; antiarthritic; antimicrobial; vulnary;  
XX cytostatic; antidiabetic; virucide; antifertility; anticonvulsant;  
XX vasotropic; antiparkinsonian; immunostimulant; dermatological;  
XX antirheumatic; antitumor; antitumor; osteopathic; tranquilizer;  
XX cerobrotective; cytokine; cell proliferation; cell differentiation;  
XX immune deficiency; severe combined immunodeficiency; SCID; tumour;  
XX autoimmune disorder; multiple sclerosis; rheumatoid arthritis;  
XX graft-versus-host disease; myeloid deficiency; wound healing; ulcer;  
XX Peridontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;  
XX Parkinson's disease; Huntington's disease; infection; cardiac disease;  
XX stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;  
XX food supplement.  
XX  
OS Homo sapiens.  
XX  
XX WO200175068-A2.  
XX  
PD 11-OCT-2001.  
XX  
XX 22-MAR-2001; 2001WO-US09369.  
XX  
XX 30-MAR-2000; 2000US-0539330.  
PR 04-DEC-2000; 2000US-0729674.  
XX  
XX (GEM ) GENETICS INST INC.  
PA

XX Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;  
PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;  
PI Clark H, Fechtel K, Merberg D;  
XX WPI: 2001-639363/73.  
DR P-PSDB; AAU39899.  
XX  
PT Secreted human proteins, useful as vaccine for treating various  
PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and  
PT nervous system disorders (e.g. stroke) -  
XX  
PS Claim 1; Page 456-457; 61pp; English.  
XX  
CC The invention relates to novel human secreted proteins, the nucleic  
CC acids encoding them. The protein may exhibit cytokine, cell proliferation  
CC or cell differentiation activity or may induce production of other  
CC cytokines in certain cell populations and may exhibit immune stimulating  
CC or immune suppressing activity, which is useful for the treatment of  
CC various immune deficiencies and disorders e.g. severe combined  
CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,  
CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary  
CC inflammation. The proteins are also useful in the treatment of diseases  
CC and disorders including tissue, skin and organ transplantation and in  
CC graft-versus-host diseases (GVHD), in the induction of tumour immunity,  
CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,  
CC in the treatment of burns, incisions and ulcers, as well as in treatment  
CC of periodontal disease, osteoporosis or osteoarthritis, mediated by  
CC inflammatory processes, diseases of the peripheral nervous system,  
CC Alzheimer's, Parkinson's disease, Huntington's disease,  
CC amyotrophic lateral sclerosis, and Shy-Drager syndrome, infections,  
CC infarction of cardiac and central nervous system vessel e.g. stroke,  
CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The  
CC protein, having activin- or inhibin-related activities is useful as a  
CC contraceptive based on the ability of inhibins to decrease fertility in  
CC female mammals and decrease spermatogenesis in male mammals. The  
CC proteins and nucleic acids are also useful as food supplements. The  
CC present sequence encodes a secreted protein of the invention.  
XX  
SQ Sequence 3871 BP; 965 A; 913 C; 1033 G; 958 T; 2 other;

Query Match 99.9%; Score 1133.2; DB 22; Length 3871;  
Best Local Similarity 100.0%; Pred. No. 9e-244;  
Matches 1134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCGGCGCAATACGACCTACAGCTCCGACGCTCCACCGGAGAGCTGATCGG 60  
DB 132 GGGCGGCGCAATACGACCTACAGCTCCGACGCTCCACCGGAGAGCTGATCGG 191

QY 61 CTCGAGTCGGCTACCGGCGCGCTGGACAGTACAGCGGCGAGCACTGGGCGGAGAGC 120  
DB 192 CTCGAGTCGGCTACCGGCGCGCTGGACAGTACAGCGGCGAGCACTGGGCGGAGAGC 251

QY 121 STKGCTACCTGGAGATGAGCTGGGCTGACCGCTGTCGCGGAGAGGAGCGCTTC 180  
DB 252 STKGCTACCTGGAGATGAGCTGGGCTGACCGCTGTCGCGGAGAGGAGCGCTTC 311

QY 181 TGCCACCGCAATGTCAGCGCGCGCGCGGCGAGCGCGCGCGGCTCGCAGCTAT 240  
DB 312 TGCCACCGCAATGTCAGCGCGCGCGCGGCGAGCGCGCGGCTCGCAGCTAT 371

QY 241 CCGGAGTCGCGCTCTTCGGGGGCTGCTGCGCGCGGCGCACTGCTCAAGCGTGGCAAG 300  
DB 372 CCGGAGTCGCGCTCTTCGGGGGCTGCTGCGCGCGGCGCACTGCTCAAGCGTGGCAAG 431

QY 301 CAGGCGCTCGGAGCTTCGCGGAGTCCGAGCCGAGCGGAGTCTGCGGAGCTTCAG 360  
DB 432 CAGGCGCTCGGAGCTTCGCGGAGTCCGAGCCGAGCGGAGTCTGCGGAGCTTCAG 491

QY 361 CCGCGGAGCGCTTACAAAGTTCCTGCGAGTTCGCTTACTTCAAGGCAATAATCTCCCAA 420  
DB 492 CCGCGGAGCGCTTACAAAGTTCCTGCGAGTTCGCTTACTTCAAGGCAATAATCTCCCAA 551

QY 421 GCCATCGCGCGCTCTCACACCTTTCTACTGAAGCATCTCTGATGACGAAATGATGAAGAGG 480  
DB 552 GCCATCGCGCGCTCTCACACCTTTCTACTGAAGCATCTCTGATGACGAAATGATGAAGAGG 611

QY 481 AACATGCGATATTATAAGAGCGCTGCTGTCGCGAGGACTACATTAAAGACCTTGAAGACC 540  
DB 612 AACATGCGATATTATAAGAGCGCTGCTGTCGCGAGGACTACATTAAAGACCTTGAAGACC 671

QY 541 AAGTCATATGAAAGCCTGTTCATCCGAGCAGTGGCGGCATACAAACGCTGAGAACTGGAGA 600  
DB 672 AAGTCATATGAAAGCCTGTTCATCCGAGCAGTGGCGGCATACAAACGCTGAGAACTGGAGA 731

QY 601 ACATCCATCACAGATGAGCTGGCGCTTCCGACTTCTTCAAGCCCTTTACGAGTGT 660  
DB 732 ACATCCATCACAGATGAGCTGGCGCTTCCGACTTCTTCAAGCCCTTTACGAGTGT 791

QY 661 CTCGCGCTTCGAGAGGTTCCAGGGAGATCAAGGACTTCAAGGATTTTCACTTTCCATA 720  
DB 792 CTCGCGCTTCGAGAGGTTCCAGGGAGATCAAGGACTTCAAGGATTTTCACTTTCCATA 851

QY 721 GCAGATCATATGTAAGTTCGGAATGCAAAATACAGTGTGAAGAGAACTTCACTTTCC 780  
DB 852 GCAGATCATATGTAAGTTCGGAATGCAAAATACAGTGTGAAGAGAACTTCACTTTCC 911

QY 781 GTTATAGGAGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTTACTTTGCA 840  
DB 912 GTTATAGGAGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTTACTTTGCA 971

QY 841 GCCTATTATAAGTTGAACGACTGAAGATGCAAGCCCTGTCAGTTCAGTTCCTGCTC 900  
DB 972 GCCTATTATAAGTTGAACGACTGAAGATGCAAGCCCTGTCAGTTCAGTTCCTGCTC 1031

QY 901 TTTGATCAGATCAGAGTTCATGACGAGAACTGGTGTATTACAGTACCAGTACCAGAGGAC 960  
DB 1032 TTTGATCAGATCAGAGTTCATGACGAGAACTGGTGTATTACAGTACCAGTACCAGAGGAC 1091

QY 961 ACTTGGGGCTCTCGGATGAGCACTTCCAGCCGAGACCTTGAAGAGTTCAGTTCTTTAAT 1020  
DB 1092 ACTTGGGGCTCTCGGATGAGCACTTCCAGCCGAGACCTTGAAGAGTTCAGTTCTTTAAT 1151

QY 1021 GTGACACACTCCAGAGAGGCTGTATGACTTTGTAAGGAAATATATGATGATGAT 1080  
DB 1152 GTGACACACTCCAGAGAGGCTGTATGACTTTGTAAGGAAATATATGATGATGATGAT 1211

QY 1081 GAGGGAGAGTTCGGAATATGTGGATGACCTTCTTGGAACTGGAGGAGACCAGC 1134  
DB 1212 GAGGGAGAGTTCGGAATATGTGGATGACCTTCTTGGAACTGGAGGAGACCAGC 1265

RESULT 3  
ABA90876  
ID ABA90876 standard; cDNA; 3871 BP.  
XX  
AC ABA90876;  
XX  
DT 14-FEB-2002 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 1.  
XX  
KW Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;  
KW immune disorder; bacterial infection; fungal infection; cancer; tumour;  
KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;  
KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;  
KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulvular;  
KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;  
KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;  
KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.  
XX  
OS Homo sapiens.  
XX  
PN US2001039335-A1.  
XX



Db 1152 GTGACCACTCCAGAGGCTGTATGACTTTGCTTAAGGAAATATATGATGATGAT 1211  
Qy 1081 GAGGGAGAGTTGTGGAATATGTGGATGACCTTTGGAACTGGAGGAGACCAGC 1134  
Db 1212 GAGGGAGAGTTGTGGAATATGTGGATGACCTTTGGAACTGGAGGAGACCAGC 1265

RESULT 4  
AAH14525  
ID AAH14525 standard; cDNA; 1961 BP.  
XX AAH14525;  
XX 25-JUN-2001 (first entry)  
DE Human cDNA sequence SEQ ID NO:12066.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX OS  
XX PN EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX PA (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX Claim 8; SEQ ID 12066; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX Sequence 1961 BP; 417 A; 579 C; 481 G; 484 T; 0 other;

Query Match 99.5%; Score 1128.4; DB 22; Length 1961;  
Best Local Similarity 99.6%; Pred. No. 8.9e-243;  
Matches 1129; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 GGGCGCGCCCAATACGAACCTACAGCTTCGGCAGCTTCCACGGGACGAGCTGATCCG 60  
Db 109 GGGCGCGCCCAATACGAACCTACAGCTTCGGCAGCTTCCACGGGACGAGCTGATCCG 168  
Qy 61 CTCGAGTCGGCTACCGCGCGCGCTGGCAAGTACAGCGCGGAGCACTGGGCGGAGAGC 120  
Db 169 CTCGAGTCGGCTACCGCGCGCGCTGGCAAGTACAGCGCGGAGCACTGGGCGGAGAGC 228  
Qy 121 STKGCTACCTGGAGATCAGCTTCGGCTGCACCGCTTGTGGCGGACGAGGCGCTTC 180  
Db 229 GTGGCTACCTGGAGATCAGCTTCGGCTGCACCGCTTGTGGCGGACGAGGCGCTTC 288  
Qy 181 TGCCACCGCAACTGCAGCGCGCGCGCGCGAGCCCGCGCGCTTCGCCAGCTAT 240  
Db 289 TGCCACCGCAACTGCAGCGCGCGCGCGCGAGCCCGCGCGCTTCGCCAGCTAT 348  
Qy 241 CCCGAGTCGGCTTCCTTCGGGGGCGCTGCTGCGCGCGCGCACTGCTCTCAAGCGCTGCAAG 300  
Db 349 CCCGAGTCGGCTTCCTTCGGGGGCGCTGCTGCGCGCGCGCACTGCTCTCAAGCGCTGCAAG 408  
Qy 301 CAGGCGCTGCCAGCTTCGCCAGTCCCGCGCGCGCGCGCGAGCGCGCGCTTCGAG 360  
Db 409 CAGGCGCTGCCAGCTTCGCCAGTCCCGCGCGCGCGCGAGCGCGCGCTTCGAG 468  
Qy 361 CGCGCGAGCGCTCAAAAGTTCTCTGAGTTCCTCAAGGCAAAATAATATCCCCAAA 420  
Db 469 CGCGCGAGCGCTCAAAAGTTCTCTGAGTTCCTCAAGGCAAAATAATATCCCCAAA 528  
Qy 421 CCCATCGCGCTGCTCAACCTTTCTACTGAAGCATCCTGATGACGAAATGATGAGAGG 480  
Db 529 CCCATCGCGCTGCTCAACCTTTCTACTGAAGCATCCTGATGACGAAATGATGAGAGG 588  
Qy 481 AACATGGCATATTATAAGAGCGCTGCTGCTGCCGAGGACTACATTAAGACCTCGAAACC 540  
Db 589 AACATGGCATATTATAAGAGCGCTGCTGCTGCCGAGGACTACATTAAGACCTCGAAACC 648  
Qy 541 AAGTCATATGAAGCGCTGTTTCATCCGAGCAGTGGGGCGCATACACGGTGAGAACTGAGA 600  
Db 649 AAGTCATATGAAGCGCTGTTTCATCCGAGCAGTGGGGCGCATACACGGTGAGAACTGAGA 708  
Qy 601 ACATCCATACAGACATGGAGCTGCGCGCTTCGGACTTCTTCAAGGCTTTTACGAGTGT 660  
Db 709 ACATCCATACAGACATGGAGCTGCGCGCTTCGGACTTCTTCAAGGCTTTTACGAGTGT 768  
Qy 661 CTCGAGCTCCGAGGGTTCCAGGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATA 720  
Db 769 CTCGAGCTCCGAGGGTTCCAGGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATA 828  
Qy 771 GCAGATCATTTATGTAGAGTTCTGGAATGCNAATACAGTGTGAAGAGACCTCACCCCA 780  
Db 839 GCAGATCATTTATGTAGAGTTCTGGAATGCNAATACAGTGTGAAGAGACCTCACCCCA 888  
Qy 781 GTTATAGAGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTTACTTGCAGTTT 840  
Db 889 GTTATAGAGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTTACTTGCAGTTT 948  
Qy 841 GCCTATTATAGTTGAACGACCTGAAGAAATGACGCCCTGTCAGCGTTCAGTATCTGCTC 900  
Db 949 GCCTATTATAGTTGAACGACCTGAAGAAATGACGCCCTGTCAGCGTTCAGTATCTGCTC 1008  
Qy 901 TTTGATCAGATGACAAAGTTCATCAGCAGAACTGTTGTTATTAACAGTACCACAGGGAC 960  
Db 1009 TTTGATCAGATGACAAAGTTCATCAGCAGAACTGTTGTTATTAACAGTACCACAGGGAC 1068  
Qy 961 ACTTGGGCGCTCTCGATGAGCACTTCCAGCCAGACCTGAAGCAGTTCAGTCTTTTAA 1020  
Db 1069 ACGTGGGCGCTCTCAGATGAGCACTTCCAGCCAGACCTGAAGCAGTTCAGTCTTTTAA 1128  
Qy 1021 GTGACCACTCCAGAGGAGCTGTATGACTTTTGTCTAAGGAAATATATGATGATGAT 1080



Db 1129 GTGACACATCCAGAGAGCTGTATGCTTTGCTAAGGAAATATAATGATGATGAT 1188  
Qy 1081 GAGGAGAGAGCTGTGGAATATGGAATGACCTCTTGGAACTGGAGAGACCCAGC 1134  
Db 1189 GAGGAGAGAGCTGTGGAATATGGAATGACCTCTTGGAACTGGAGAGACCCAGC 1242

## RESULT 5

AAH311138  
ID AAH311138 standard; cDNA; 1586 BP.

AC AAH311138;

XX 27-JUL-2001 (first entry)

XX Human colon cancer cell line Kml2L4-A cDNA library derived seq#1072.

XX Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;

KW detection; colon cancer cell line Kml2L4-A; ss.

XX Homo sapiens.

OS W0200018916-A2.

PN 06-APR-2000.

XX 23-SEP-1999; 99WO-US22226.

XX 28-SEP-1998; 98US-0102151.

PR 28-SEP-1998; 98US-0102150.

PR 29-SEP-1998; 98US-0102390.

PR 08-OCT-1998; 98US-0103815.

PR 27-OCT-1998; 98US-0105877.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;

XX WPI; 2000-293155/25.

XX Polynucleotide library comprising 1079 defined sequences, useful in

PT the form of an array to detect cancer or susceptibility to cancer -

XX Claim 1; Page 498-499; 502pp; English.

XX The present invention describes a library of polynucleotides comprising  
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described  
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to  
CC one of the 1079 sequences; (2) a recombinant host cell containing (1);  
CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that  
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method  
CC of detecting differentially expressed genes correlated with a cancerous  
CC state of a mammalian cell comprising detecting a gene product encoded by  
CC one of the 1079 sequences given in the specification. The polynucleotides  
CC are used to monitor patients having (or susceptible) to cancer to detect  
CC potentially malignant events at a molecular level before they are  
CC detectable at a gross morphological level. The polynucleotides are also  
CC useful for monitoring the efficacy of various therapies and preventive  
CC interventions. Polynucleotide probes based on the disclosed sequences  
CC are useful for chromosome mapping and detection of transcription levels.  
CC The 1079 polynucleotide sequences were derived from a human colon cancer  
CC cell line Kml2L4-A cDNA library.

XX Sequence 1586 BP; 381 A; 406 C; 340 G; 453 T; 6 other;

XX Query Match 62.2%; Score 705; DB 21; Length 1586;

XX Best Local Similarity 99.3%; Pred No. 4,4e-148;

XX Matches 702; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 428 CCGCTGCTCACACCTTTCTACTGAAGCATCCTGATGACGAATGATGAAGAGAAACATGG 487  
Db 1 CCGCTGCTCACACCTTTCTACTGAAGCATCCTGATGACGAATGATGAAGAGAAACATGG 60  
Qy 488 CATATTATAGAGAGCTGCTGGTCCGAGGACTACATTAAAGACCTTGAACCAAGTCAT 547  
Db 61 CATATTATAGAGAGCTGCTGGTCCGAGGACTACATTAAAGACCTTGAACCAAGTCAT 120  
Qy 548 ATGAAGCCTGTTTCATCCGAGCAGTGGGGCATCAACAGGTGAGAACTGAGAGAAACATCCA 607  
Db 121 ATGAAGCCTGTTTCATCCGAGCAGTGGGGCATCAACAGGTGAGAACTGAGAGAAACATCCA 180  
Qy 608 TCACAGACATGAGAGCTGCTGGTCCGAGGACTTCCGAGCTTCTTCAAAGCCTTTTACGAGTGTCTCCAG 667  
Db 181 TCACAGACATGAGAGCTGCTGGTCCGAGGACTTCTTCAAAGCCTTTTACGAGTGTCTCCAG 240  
Qy 668 CCGTGGAGGCTTCCAGGAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAGATC 727  
Db 241 CCGTGGAGGCTTCCAGGAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAGATC 300  
Qy 728 ATTATGTAGAAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAACTCACCCAGTTATAG 787  
Db 301 ATTATGTAGAAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAACTCACCCAGTTATAG 360  
Qy 788 GAGGCTATCCGTTGAGAAATTTCTGGCTACCATGATCATTTACTTGCAGTTTGCCTATT 847  
Db 361 GAGGCTATCCGTTGAGAAATTTCTGGCTACCATGATCATTTACTTGCAGTTTGCCTATT 420  
Qy 848 ATAAGTTGAAGAGCTGGAAGAAATGAGCCCTCTGAGTCAGTATCTGCTCTTTGATC 907  
Db 421 ATAAGTTGAAGAGCTGGAAGAAATGAGCCCTCTGAGTCAGTATCTGCTCTTTGATC 480  
Qy 908 AGAATGACAAGGTGATGAGCAGAGAACTGGTGTATTACCATGACACAGGAGACATGGG 967  
Db 481 AGAATGACAAGGTGATGAGCAGAGAACTGGTGTATTACCATGACACAGGAGACATGGG 540  
Qy 968 GCCTCTCGGATGAGCAGCTTCCAGCCAGACCTGGAAGCAGTTCACTTTTAAATGTGACCA 1027  
Db 541 GCCTCTCGGATGAGCAGCTTCCAGCCAGACCTGGAAGCAGTTCACTTTTAAATGTGACCA 600  
Qy 1028 CACTCCAGAGAGGCTGATGACCTTTGCTAAGGAAATATTAATGATGATGATGAGGAG 1087  
Db 601 CACTCCAGAGAGGCTGATGACCTTTGCTAAGGAAATATTAATGATGATGATGAGGAG 660  
Qy 1088 AAGTTGTGGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1134  
Db 661 AAGTTGTGGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 707

## RESULT 6

AAH06808

ID AAH06808 standard; cDNA; 818 BP.

XX AAH06808;

XX 26-JUN-2001 (first entry)

DE Human cDNA clone (5'-primer) SEQ ID NO:3643.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.  
PR 05-JUN-2000; 2000JP-0241899.  
PA (HELI-) HELIX RES INST.  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 1; SEQ ID 3643; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC polynucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 818 BP; 159 A; 271 C; 236 G; 148 T; 4 other;  
  
Query Match 51.8%; Score 587.6; DB 22; Length 818;  
Best Local Similarity 96.2%; Pred. No. 6.7e-122;  
Matches 631; Conservative 2; Mismatches 18; Indels 5; Gaps 3;  
  
QY 1 GGGCGCGCCCATACGAAGCTACAGCTTCCCGAGCTTCCACGGGAGAGCTGATGCCG 60  
DB 109 GGGCGCGCCCATACGAAGCTACAGCTTCCCGAGCTTCCACGGGAGAGCTGATGCCG 168  
QY 61 CTCGAGTCGGCTTACCGGACCGCTGGCAAGTACAGCGGAGCACTGGGCCGAGAGC 120  
DB 169 CTCGAGTCGGCTTACCGGACCGCTGGCAAGTACAGCGGAGCACTGGGCCGAGAGC 228  
QY 121 SYKGGCTACTCGAGATCAGCTGCGGTGACCGCTTCTGCGGACGAGGAGCCCTTC 180  
DB 229 GTGGGCTACTCGAGATCAGCTGCGGTGACCGCTTCTGCGGACGAGGAGCCCTTC 288  
QY 181 TGCCACCGCAACTGCGAGCCGCGCGCGGAGCCGAGCCGCGCGCTGCCAGCTAT 240  
DB 289 TGCCACCGCAACTGCGAGCCGCGCGCGGAGCCGAGCCGCGCGCTGCCAGCTAT 348  
QY 241 CCCGAGCTCGGCTCTTTCGGGGGCTGCTGTCGCCCGGCGCACTGCTCAAGCGTGCAG 300  
DB 349 CCCGAGCTCGGCTCTTTCGGGGGCTGCTGTCGCCCGGCGCACTGCTCAAGCGTGCAG 408  
QY 301 CAGGGCTGCCAGCTTCCGCGAGTCCAGCCAGCCGCGAGGTCTCGGCGGACTTCGAG 360  
DB 409 CAGGGCTGCCAGCTTCCGCGAGTCCAGCCAGCCGCGAGGTCTCGGCGGACTTCGAG 468  
QY 361 CCGCGCGAGCCCTACAAGTTCCTCGAGTTCGCTTACTTCAAGGCAATAATCTCCCAAAA 420  
DB 469 CCGCGCGAGCCCTACAAGTTCCTCGAGTTCGCTTACTTCAAGGCAATAATCTCCCAAAA 528

QY 421 GCCATGCCCGCTGCTCACACCTTTTACTGAAGCATCCTGATGACCAAAATGATGAAGAGG 480  
DB 529 GCCATGCCCGCTGCTCACACCTTTTACTGAAGCATCCTGATGACCAAAATGATGAAGAGG 588  
QY 481 AACATGCGATATTATAAGAGCCCTGCTGTCGCGAGGACTACATTAAAGACCTTGGAAACC 540  
DB 589 AACATGCGATATTATAAGAGCCCTG-CTGCTGCGGAGGACTACATTAAAGACCTTGGAAACC 647  
QY 541 AAGT--CATATGAAAGCCTGTTTCATCCGAGCAGTGGCGGCATACAAACGGTGAGAACTGGA 598  
DB 648 AAGTTCATATGAAAGCCTGTTTCATCCGANCAGTGGCGGCATACAAACGGTGAGAACTGGA 707  
QY 599 GAACATCCATCAGACAGATGAGAGCTGGCCC--TTCCGAGCTTCTTCAAAGCCCTTTT 652  
DB 708 GAACATTCATCAGACAGATGAGAACTGGCCCTTTCCGGAATTTCTTCAAAAGNCTTT 763  
  
RESULT 7  
ABL67662  
ID ABL67662 standard; DNA; 2347 BP.  
XX  
AC ABL67662;  
XX  
DT 15-MAY-2002 (first entry)  
DX  
DE Oesophagus cancer related gene sequence SEQ ID NO:5999.  
XX  
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
FN WO200194629-A2.  
XX  
PD 13-DEC-2001.  
XX  
PP 30-MAY-2001; 2001WO-US10838.  
XX  
PR 05-JUN-2000; 2000US-209473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 25-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 27-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.  
PR 28-SEP-2000; 2000US-236033P.  
PR 28-SEP-2000; 2000US-236034P.  
PR 28-SEP-2000; 2000US-236109P.  
PR 28-SEP-2000; 2000US-236111P.  
PR 29-SEP-2000; 2000US-236842P.  
PR 29-SEP-2000; 2000US-236891P.  
PR 02-OCT-2000; 2000US-237172P.  
PR 02-OCT-2000; 2000US-237173P.  
PR 02-OCT-2000; 2000US-237278P.

PR 02-OCT-2000; 2000US-237294P.  
PR 02-OCT-2000; 2000US-237295P.  
PR 02-OCT-2000; 2000US-237316P.  
PR 03-OCT-2000; 2000US-237425P.  
PR 03-OCT-2000; 2000US-237598P.  
PR 03-OCT-2000; 2000US-237604P.  
PR 03-OCT-2000; 2000US-237606P.  
PR 03-OCT-2000; 2000US-237608P.  
PR 01-NOV-2000; 2000US-244857P.  
PR 01-NOV-2000; 2000US-245084P.  
XX  
PA (AVAL-) AVALON PHARM.  
XX  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX WPI; 2002-188264/24.  
XX  
XX Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set  
XX  
XX Claim 1; SEQ ID 5999; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABLe1664  
CC to ABLe70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.  
XX  
SQ Sequence 2347 BP; 517 A; 689 C; 733 G; 408 T; 0 other;

Query Match. 40.2%; Score 455.4; DB 24; Length 2347;  
Best Local Similarity 65.1%; Pred. No. 3.3e-92;  
Matches 723; Conservative 2; Mismatches 368; Indels 18; Gaps 3;

QY 5 GCGGCCAATACGACGCTACAGCTTCCGAGCTTCCACCGGACGAGCTGATGCCGCTCG 64  
DB 61 GGGCGCAGTACGAGNAGTACAGCTTCCGGGCTTCCGCCCGGAGACCTGATGCCGCTGG 120  
QY 65 AGTCGGCTACCGGACGCGCTGAGCAAGTACAGCGCGGAGAGCTGGCGGAGAGGCTKG 124  
DB 121 CCGGGCGGTACGGGACGCTCTGGAGCAGTACGAGGAGAGAGCTGGCGGAGAGCGCG 180  
QY 125 GCTACCTGGAGATCAGCTCGGCTGACCGCTTGTCTGCGGACGAGGCGCTTCTGCC 184  
DB 181 GCTACCTGGAGGCGGCGCTCGGCTGCACCGCTTCTGCGGACGAGGCGCTTCTGCC 240  
QY 185 ACCGCAACTCGAGCG-----CCGCGCGGACGCGGCGGCGCGGCGCTCGCCA 235  
DB 241 AGGCCAACTCAGCGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
QY 236 GCTA-----TCCGAGCTGGGCTCTTCCGGGCGCTGCTCGCGGCGGCGGCGCTCA 289  
DB 301 ACAGTGGGCTGGAGCTGGGCTCTTCCGGGCGGCTCTGAGGCGGCGGCGGCGGCTGC 360  
QY 290 AGCGCTGCAAGCGGCTGCGAGCTTCCGCCAGTCCCGAGCGGCGGCGGAGGCTGCG 349  
DB 361 GGGCTGCAAGCGGACGCTGCGGCTTCCAGGCTTCCAGGCTGCGCTACCGCGGCGGCGGCTGCG 420  
QY 350 CGGACTTCCAGCGCGGCGGAGCGCTTACAGTCTCTGCGAGTTCGCTTACAGGCAATA 409

## RESULT 8

AAH30585  
ID AAH30585 standard; cDNA; 408 BP.

XX AAH30585;

XX 27-JUL-2001 (first entry)

XX Human colon cancer cell line Kml2L4-A cDNA library derived sequence #519.  
XX Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;  
KW detection; colon cancer cell line Kml2L4-A; ss.

XX Homo sapiens.

XX W0200018916-A2.

XX 06-APR-2000.

XX 23-SEP-1999; 99WO-US22226.

PR 28-SEP-1998; 98US-0102161.  
PR 28-SEP-1998; 98US-0102180.  
PR 29-SEP-1998; 98US-0102380.  
PR 08-OCT-1998; 98US-0103815.  
PR 27-OCT-1998; 98US-0105877.  
XX  
XX (CHIR ) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
XX  
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
PI Reinhard C, Giese K, Randazzo P, Kennedy GC, Pot D, Kassam A;  
PI Lanson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
XX  
XX WPT; 2000-293155/25.  
XX  
XX Polynucleotide library comprising 1079 defined sequences, useful in  
PT the form of an array to detect cancer or susceptibility to cancer -  
XX  
XX Claim 1; Page 333; 502pp; English.  
XX  
XX The present invention describes a library of polynucleotides comprising  
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described  
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to  
CC one of the 1079 sequences; (2) a recombinant host cell containing (1);  
CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that  
CC specifically binds to (II); (5) a vector comprising (1); and (6) a method  
CC of detecting differentially expressed genes correlated with a cancerous  
CC state of a mammalian cell comprising detecting a gene product encoded by  
CC 65 of the 1079 sequences given in the specification. The polynucleotides  
CC are used to monitor patients having (or susceptible) to cancer to detect  
CC potentially malignant events at a molecular level before they are  
CC detectable at a gross morphological level. The polynucleotides are also  
CC useful for monitoring the efficacy of various therapies and preventive  
CC interventions. Polynucleotide probes based on the disclosed sequences  
CC are useful for chromosome mapping and detection of transcription levels.  
CC The 1079 polynucleotide sequences were derived from a human colon cancer  
CC cell line Kml2L4-A cDNA library.  
XX  
SQ Sequence 408 BP; 117 A; 96 C; 98 G; 97 T; 0 other;

Query Match 35.8%; Score 406.4; DB 21; Length 408;  
Best Local Similarity 95.8%; Pred. No. 1.9e-81;  
Matches 407; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 428 CCGCTGCTCACACCTTCTACTGAAGCATCTGATGACGAATGATGAGGAGACATGG 487  
Db 1 CCGCTGCTCACACCTTCTACTGAAGCATCTGATGACGAATGATGAGGAGACATGG 60  
Qy 488 CATATTATAAGAGCTGCTGCTGGTCCGAGGACTACATTAAAGACCTGGAACCAAGTCAT 547  
Db 61 CATATTATAAGAGCTGCTGCTGGTCCGAGGACTACATTAAAGACCTGGAACCAAGTCAT 120  
Qy 548 ATGAAGCCTGTTTCATCCGAGCAGTGGGGCATACACGGTGAGACTGAGACATCCA 607  
Db 121 ATGAAGCCTGTTTCATCCGAGCAGTGGGGCATACACGGTGAGACTGAGACATCCA 180  
Qy 608 TCACAGACATGGAGCTGGCCCTTCCCGACCTTCTCAAAGCCTTTTACGAGTGTCTGCGAG 667  
Db 181 TCACAGACATGGAGCTGGCCCTTCCCGACCTTCTCAAAGCCTTTTACGAGTGTCTGCGAG 240  
Qy 668 CCTCGAGGGTTTCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATGACGATC 727  
Db 241 CCTCGAGGGTTTCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATGACGATC 300  
Qy 728 ATTATGTAGAGTTCTGMAATGCAAAATACAGTGTGAAGAGACCTCACCCAGTTATAG 787  
Db 301 ATTATGTAGAGTTCTGMAATGCAAAATACAGTGTGAAGAGACCTCACCCAGTTATAG 360  
Qy 788 GAGGCTATCCGGTTGAGAAATTTGGCTACCATGTATCTATTCTTC 835  
Db 361 GAGGCTATCCGGTTGAGAAATTTGGCTACCATGTATCTATTCTTC 408

RESULT 9  
AB211344  
ID AB211344 standard; cDNA; 1309 BP.  
XX  
XX AC AB211344;  
XX  
XX DT 20-JAN-2003 (first entry)  
XX  
XX Human polynucleotide SEQ ID NO 226.  
XX  
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cystostatic; immunomodulator; nootropic; neuroprotective;  
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
KW antiarthritic; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX WC200270539-A2.  
XX  
XX 12-SEP-2002.  
PD  
XX 05-MAR-2002; 2002WO-US05095.  
PF  
XX 05-MAR-2001; 2001US-0799451.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac R;  
XX  
XX WPI; 2002-759812/82.  
DR  
XX P-PSDB; ABP69127.  
XX  
XX New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
PT platelet or coagulation disorders -  
XX  
XX Claim 1; SEQ ID NO 226; 1012pp + Sequence Listing; English.  
PS  
XX The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences  
CC (AB211119-AB212066) or their mature protein coding portion, active domain  
CC coding protein or complementary sequences. The polynucleotides are useful  
CC for identifying expressed genes or for physical mapping of human genome.  
CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular  
CC weight markers, as a food supplement, for generating antibodies, in  
CC medical imaging, screening and diagnostic assays and for treating  
CC cell-proliferative disorders (cancer), neurodegenerative diseases  
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
CC disorders, platelet or coagulation disorders, wound, burns, incision,  
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
CC parasitic), arthritis, etc.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1309 BP; 287 A; 384 C; 365 G; 273 T; 0 other;

Query Match 13.8%; Score 156.4; DB 24; Length 1309;  
Best Local Similarity 53.6%; Pred. No. 2.2e-25;  
Matches 437; Conservative 0; Mismatches 346; Indels 33; Gaps 4;  
Qy 153 CCGCTGTGTCGCGACAGCGAGGCTTCTGCCACCGCAACTGCAGCGCGCGCGACCC 212  
Db 256 CGCACCCAGTGTGCGCGGACTTCCGCTGGAGTGCACCCCGACTGGTCCCCCGACCC 315

QY 213 CGAGCCCGCCGCGGCTCGCAGCTATCCGAGTGGCGCTCTTCGGGGGCTGCTGCG 272  
 DB 316 GGGCCAGGCTCGGGCGCGCGGCTCGGAGCTTCTTCGGGGGCTTCTGCG 375  
 QY 273 CGGCGGCACTGCTCAAGCGCTGACAGCGGCTGCGAGCTTCCGAGTCCCGAGCC 332  
 DB 376 TCGGCTGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 426  
 QY 333 CAGCGCGAGTGTGCGGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 392  
 DB 427 CAGCGAGAGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 486  
 QY 393 TTACTTCAAGCAATAATCTCCCAAGCAATGCGGCTGCTGCGAGCTTCTTACTGAA 452  
 DB 487 CTACTTCAAGATCAACAAGTTGGAGAAAGCTGTTCTGCGAGCACACACCTTCTTCTG 546  
 QY 453 GCATCTCATGACGAATGATGAGAGAAATGATGAGAGAAATGATGAGAGAAATGATGAG 511  
 DB 547 CAATCTGAGCAGATGAGAAATGAGAGAAATGATGAGAGAAATGATGAGAGAAATGATGAG 606  
 QY 512 --CGAGGATACATTAAGACCTGGAACCAAGTATATGAAGACCTGTTATCGAGC 569  
 DB 607 GAAGAGGCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGG 666  
 QY 570 AGTGGGCTATACAGGCTGAGACTGAGAGAAATGATGAGAGAAATGATGAGAGAAATGATGAG 629  
 DB 667 AGTGGGCTATACAGGCTGAGAAATGAGAGAAATGATGAGAGAAATGATGAGAGAAATGATGAG 726  
 QY 630 TCCGAGCTTCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCT 689  
 DB 727 GCAAGATACATTTGCGGCTATGAGGAGTGGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 786  
 QY 690 CAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGG 731  
 DB 787 CGATGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGG 846  
 QY 732 TGTGAAGTCTGGAATCAAAATACAGTGTGGAAGAACTTCAAGGCTTCAAGGCTTCAAGGCT 783  
 DB 847 CATCAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGG 906  
 QY 789 AGGCTATCGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGG 848  
 DB 907 AGAGAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGG 966  
 QY 849 TAAGTTGAAGCCTGAGAGATGAGAGAAATGATGAGAGAAATGATGAGAGAAATGATGAGAGAAAT 908  
 DB 967 TAACATGGAGATATACAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGG 1026  
 QY 909 GAATGACAGGCTATGAGAGAAATGATGAGAGAAATGATGAGAGAAATGATGAGAGAAATGATGAGAGAAAT 944  
 DB 1027 CAATGACAGGCTATGAGAGAAATGATGAGAGAAATGATGAGAGAAATGATGAGAGAAATGATGAGAGAAAT 1062

RESULT 10  
 ID ABZ11345  
 XX ABZ11345 standard; cDNA; 2152 BP.  
 AC ABZ11345;  
 XX  
 DT 20-JAN-2003 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 227.  
 KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease; multiple  
 KW sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 KW arthritis; cyclostatic; immunomodulator; nontropic; neuroprotective;  
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
 KW antiarthritic; gene; ss.

OS Homo sapiens.  
 XX  
 PN WO200270539-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 05-MAR-2002; 2002WO-US05095.  
 XX  
 PR 05-MAR-2001; 2001US-0799451.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX  
 DR WPI; 2002-759812/82.  
 DR  
 DR P-PSDB; ABP69126.  
 XX  
 PT New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative,  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
 PT platelet or coagulation disorders  
 XX  
 PS Claim 1; SEQ ID NO 227; 1012pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (I) comprising a  
 CC nucleotide sequence selected from any of 948 sequences  
 CC (AB21119-AB212066) or their mature protein coding portion, active domain  
 CC coding protein or complementary sequences. The polynucleotides are useful  
 CC for identifying expressed genes or for physical mapping of human genome.  
 CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular  
 CC weight markers, as a food supplement, for generating antibodies, in  
 CC medical imaging, screening and diagnostic assays and for treating  
 CC cell-proliferative disorders (cancer), neurodegenerative diseases  
 CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
 CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
 CC disorders, platelet or coagulation disorders, wound, burns, incision,  
 CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
 CC parasitic), arthritis, etc.  
 CC Note: the sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2152 BP; 494 A; 636 C; 602 G; 420 T; 0 other;

Query Match 13.8%; Score 156.4; DB 24; Length 2152;  
 Best Local Similarity 53.6%; Pred. No. 2.5e-25;  
 Matches 437; Conservative 0; Mismatches 346; Indels 33; Gaps 4;  
 QY 153 CCGCTTGTCTGCGGAGAGGCTTCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCT 212  
 DB 256 CCGCAGGCTGCGGAGCTTCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCT 315  
 QY 213 CGAGCCCGCGGCTGCGGAGCTATCCGAGTGGCGCTTCTTCGGGGGCTGCTGCG 272  
 DB 316 GGGCCAGGCTCGGGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 375  
 QY 273 CCGCGGCACTGCTCAAGCGCTGCAAGCGCTGCAAGCGCTGCAAGCGCTGCAAGCGCTGCAAGCG 332  
 DB 376 TCGGCTGCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 426  
 QY 333 CAGCGCGAGTGTGCGGAGCTTCCAGCGCGGAGGCTTCCAGGCTTCTGCGAGTTCGCTGCG 392  
 DB 427 CAGCGAGAGATGAGCTGAGCTGAGCTTCCGCAAGCGGAGCCCTTCAACTACCTGCGAGTTCG 486  
 QY 393 TTACTTCAAGCAATAATCTCCCAAGCAATGCGGCTGCTGCGAGCTTCTTACTGAA 452  
 DB 487 CTACTTCAAGATCAACAAGTTGGAGAAAGCTGTTCTGCGAGCACACACCTTCTTCTG 546  
 QY 453 GCATCTCATGACGAATGATGAGAGAAATGATGAGAGAAATGATGAGAGAAATGATGAGAGAAAT 511  
 DB 547 CAATCTGAGCAGATGGAATGAGAGAAATGAGAGAAATGATGAGAGAAATGATGAGAGAAATGATGAG 606



Db 816 CATCCAGTCTCAACTTAAGCAGAACTGTGTACAGGAGTGTCTCCCAACCAAGTCG 875  
Qy 789 AGGCTATCCGGTTGAGAAATTTGGCTACCAATGATATCACTTACGAGTTGGCTATTA 848  
Db 876 AGAGAGCCCTTTGAAGACTTCTCCCATCGCAATTAATATCTGCAAGTTGGCTACTA 935  
Qy 849 TAAGTTGAACGCTGAGAAAGCAGCCCTGTGCACTGAGTATCTGCTTTGATCA 908  
Db 936 TAACATTTGGAAATTAACACAGGCTGTGAATGTGCCAAGACCTATCTCTCTTCCC 995  
Qy 909 GAATGACAAAGGTCATCGACAGCAACCTGTGTATTA 944  
Db 996 CATGACAGGCTGATGACCAAAATTTGGCTATTA 1031

## RESULT 12

ABZ11346

ID ABZ11346 standard; cDNA; 2242 BP.

XX

AC ABZ11346;

XX

DT 20-JAN-2003 (first entry)

XX

DE Human polynucleotide SEQ ID NO 228.

XX

KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cystic; immunomodulator; nontropic; neuroprotective;  
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
KW antiarthritic; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200270539-A2.

XX

PD 12-SEP-2002.

XX

XX 05-MAR-2002; 2002WO-US05095.

XX

XX 05-MAR-2001; 2001US-0799451.

PR

XX (HYSE-) HYSEQ INC.

XX

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;

XX

DR WPI; 2002-759812/82.

DR

PS P-PSDB; ABP69129.

XX

PT New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
PT platelet or coagulation disorders

XX

PS Claim 1; SEQ ID NO 228; 1012pp + Sequence Listing; English.

XX

CC The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences  
CC (ABZ1119-ABZ12086) or their mature protein coding portion, active domain  
CC coding protein or complementary sequences. The polynucleotides are useful  
CC for identifying expressed genes or for physical mapping of human genome.

CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular  
CC weight markers, as a food supplement, for generating antibodies, in  
CC medical imaging, screening and diagnostic assays and for treating  
CC cell-proliferative disorders (cancer), neurodegenerative diseases  
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
CC disorders, platelet or coagulation disorders, wound, burns, incision,  
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,

CC parasitic), arthritis, etc.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 2242 BP; 518 A; 655 C; 625 G; 444 T; 0 other;

Query Match 13.8%; Score 156.4; DB 24; Length 2242;

Best Local Similarity 53.6%; Pred. No. 2.5e-25;

Matches 437; Conservative 0; Mismatches 346; Indels 33; Gaps 4;

Qy 153 CGCTTGTCTGCGGACAGCAGGCGCTTCTGCCACCGCACTCAGCGCGCGCAGCC 212

Db 256 CGCACCCAGTGTGCGCGGCGCTTCCGTGGAGTGGACCCGACTGTGTCCCGAGCCC 315

Qy 213 CGAGCCCCGCGCGGCTCTGCCAGCTATCCCGAGTGTGCGCTTCTGGGGGCTGCTGG 272

Db 316 GGCCAGGCTCGGGCGCGCGCTGCGGACCTGAGCTTCTTGGGGGCTTCTGGG 375

Qy 273 CGCGGCGCACTGCTCAAGCGCTGCAAGCAGGCGCTGCCAGCTTCCGCCATGCCAGCC 332

Db 376 TCGCGTGTCTGCTGCGCGGCTGCGCTGCGGCGCGCGCGCGCGCTC-----GCT 426

Qy 333 CAGCCGCGAGTGTCTGCGGAGCTTCCAGCGCGCGCGCGCGCTACAAAGTTCTCTCAGTTCG 392

Db 427 CAGCGAAGAGTGTGAGCTGAGTTCCGCAAGCGGAGCGCTTACAACTACTCTCAGTTCG 486

Qy 393 TTACTTCAAGGCAATAATCTCCCAAGCCATCGCGCTGTCTACACCTTCTACTGAA 452

Db 487 CTACTTCAAGATCAACAAGTTGAGAAAGCTGTGCTGCAGCACACACCTTCTTCTGGG 546

Qy 453 GCATCTGTATGACAAATGATGAAGAGGAACATGGCATATTAAGAGCTCTGCTGGTG- 511

Db 547 CATCTGTGACATGGAATGCAGCAGAACTAGACTATTACCAACCATCTCTGGAGT 606

Qy 512 --CCGAGGACTACATTAAGACCTGGAACACCAAGTATATGAAGCTTCTCATCCGAGC 569

Db 607 GAAGGAGGCGGAGCTTCAAGGATCTTGAGACTCAACCCATATGCAAGAAATTCGACTGGG 666

Qy 570 AGTGGCGGCATACAAGCGTGAGAACTGGAGAACATCCATCAGACATGAGCTGCGCCT 629

Db 667 AGTGGCACTTCTAGAGGAACAGCCACAGAAAGTGTGCGCCCTAGAGCGCGGCT 726

Qy 630 TCCGAGCTTCTTCAAGACCTTTTACAGTGTCTCGCAGCTCGAGGTTCCAGGGAGAT 689

Db 727 GCAAGAATACTTTGTGGCTATGAGAGTGTGCTGCTGCAAGGCGCTTATGACTA 786

Qy 690 CAGGAGCTTCAA-----GGATTCTACCTTTCATAGCAGATCATTA 731

Db 787 CGATGCTCAACTACCTTTGAGTACAAACGCTGACCTTTCAGGCGCATCACAGATCATTA 846

Qy 732 TGTAGAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAACCT---CACCCAGTTATAGG 788

Db 847 CATCCAGTCTCTCACTGTAAAGCAGAACTGTGTACGAGCTTGTCTCCACCCAGTCG 906

Qy 789 AGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTTCTGAGTTTGCCTATTA 848

Db 907 AGAGAAGCCCTTTGAGACTTCTCCCATCGCAATTAATTAATCTGCAAGTTGCTTACTA 966

Qy 849 TAAGTTGAACGCTGGAAGAAATGACGCGCTGTGAGTGTGAGTGTCTCTTTGATCA 908

Db 967 TAACATTTGGAATTAACACAGCTGTGGAATGTGCCAAGACCTATCTCTCTTCTCC 1026

Qy 909 GAATGCAAGGCTCATGACAGCAAACTGGTGATTA 944

Db 1027 CAATGACGAGGTGATGAACCAAAATTTGGCTTATTA 1062

## RESULT 13

AAH14513

ID AAH14513 standard; cDNA; 2524 BP.

XX

AC AAH14513;



XX 26-JUN-2001 (first entry)  
 XX Human cDNA sequence SEQ ID NO:12044.  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX Homo sapiens.  
 XX EP1074617-A2.  
 XX 07-FEB-2001.  
 XX 28-JUL-2000; 2000EP-0116126.  
 XX 29-JUL-1999; 99JP-0248036.  
 XX 27-AUG-1999; 99JP-0300253.  
 XX 11-JAN-2000; 2000JP-0118776.  
 XX 02-MAY-2000; 2000JP-0183767.  
 XX 09-JUN-2000; 2000JP-0241899.  
 XX (HELI-) HELIX RES INST.  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI; 2001-318749/34.  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX Claim 8; SEQ ID 12044; 2537pp + CD ROM; English.  
 XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH2446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 2524 BP; 583 A; 747 C; 704 G; 490 T; 0 other;  
 Query Match 13.8%; Score 156.4; DB 22; Length 2524;  
 Best Local Similarity 53.6%; Pred. No. 2.6e-25;  
 Matches 437; Conservative 0; Mismatches 346; Indels 33; Gaps 4;  
 QY 153 CGCTTGTGCGGACAGGAGGCTTCTGCGACCGCACTGACGCGCGCGCGCC 212  
 DB 240 CGCACCAAGTGTGCGCGGACCTTCCGTGGAGGTGACCCCGACTGTCGCCAGCCC 299  
 QY 213 CGAGCCCGCGCGGCTCGCCAGCTATCCGAGCTGCGCTCTTCGGGGGCGCTGCTGCG 272  
 DB 300 GCGCCAGGCGCTCGGGGCGCGCGCCCTGCGACCTGAGCTTCTTCGGGGGCGCTTCTGCG 359  
 QY 273 CCGCGGCGACTGCCTCAAGCGCTGCGAGCAGGCGCTGCCAGCTTCGCCAGTCCAGCC 332

DB 360 TCGGCTGCTGCTGCGCGGCTGCTGCGGCGCGCGCGCGCGCTC-----GCT 410  
 QY 333 CAGCGCGAGGTGCTGCGGACTTCCAGCGCGCGGAGCCCTACAAGTTCCTGCGAGTTCG 392  
 DB 411 CAGGAAGAGTGGAGCTGGAGTTCCGGAAGCGAGCCCTACAACCTACCTGCGAGTTCG 470  
 QY 393 TTACTTCAAGGCAATAATCTCCCAAGCCATCGCGCTGCTCACAACCTTTCTACTGAA 452  
 DB 471 CTACTTCAAGATCAACAAGTTGGAGAAAGCTGTTGTCGAGCACACACCTTCTTCGTGG 530  
 QY 453 GCATCTGATGACGAATGATGAAGAGACATGGCATATTATAGAAGCTGCTGCTG- 511  
 DB 531 CAATCTCTGAGCACATGGAATGCGAGAGAACCTAGACTATTACCAAAACCATGTCGAGT 590  
 QY 512 --CGAGGACTACATTAAAGACCTGGAACCAAGTCAATATGAAAGCTGTTCATCGGAGC 569  
 DB 591 GAAGGAGCGGACTTCAAGGATCTTGAGACTCAACCCATATGCAAGAATTTTCGACTGG 650  
 QY 570 AGTGGGGCATACAACGGTGAAGAACTGGAGAACATCCATCACAAGATGAGGCTGCCCC 629  
 DB 651 AGTGGCACTTACTCAGAGGACACAGCCACAGGAAGCTGTGCCCCACCTAGAGCGCGCT 710  
 QY 630 TCCGACTTCTTCAAAGCCCTTTTACGAGTGTCTCGCAGCTGCGAGGTTCCAGGAGAT 689  
 DB 711 GCAAGAAATCTTTGTGCGCTATGAGGAGTCCCTGCGAAGGCGCTTATGACTA 770  
 QY 690 CAAGGACTTCAA-----GGATTCTTACCTTTCCATAGCAGATCATTA 731  
 DB 771 CGATGGCTACAACCTACCTTGAGTACAACGCTGACCTTCCAGGCCATCACAGATCATTA 830  
 QY 732 TGTAAGATTCTGGAATGCAAAATACAGTGTGAAGAACCT---CACCCAGTTATAG 788  
 DB 831 CATCCAGTCTCTCACTGTAAAGCAGAACTGTGTCAGGAGCTTGCTTCCACCCCAAGTCG 890  
 QY 789 AGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTTACTTGCAGTTTGCCTATTA 848  
 DB 891 AGAAGAGCCCTTTGAAGACTTCTCCCATCGCATTAATAATTATCTGCAGTTTGCCTACTA 950  
 QY 849 TAAGTTGAAGCACTGAAGAAATGCAAGTCCAGCCCTGTCGAGTCAGCTATCTGCTTTGATCA 908  
 DB 951 TAACTTGGGAATTATACACAGCTGTTGAATGTGCCAAGACCTATCTTCTCTCTTCCC 1010  
 QY 909 GAATGACAGGTCATGACGAGCAACCTGGTGTATTA 944  
 DB 1011 CAATGACAGGTCATGAACCAAAATTTGGCTATTA 1046  
 RESULT 14  
 AAF93800  
 ID AAF93800 standard; cDNA; 2563 BP.  
 AC AAF93800;  
 XX  
 DT 23-MAY-2001 (first entry)  
 XX Human cDNA encoding a membrane or secretory protein clone PSEC0109.  
 DE Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KW rheumatoid arthritis; diabetes; ss.  
 XX Homo sapiens.  
 XX EPI067182-A2.  
 XX 10-JAN-2001.  
 XX 07-JUL-2000; 2000EP-0114090.  
 XX 08-JUL-1999; 99JP-0194179.  
 PR 11-JAN-2000; 2000JP-0118775.  
 PR 02-MAY-2000; 2000JP-0183766.  
 XX

PA (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
XX WPI: 2001-092989/11.  
XX P-PSDB; AAB88373.  
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in  
XX gene therapy or as candidate target molecules in drug development -  
XX Claim 1; SEQ ID 113; 609pp + CD ROM; English.  
XX This invention relates to nucleic acid sequences AAF93744 - AAF93916  
XX which encode human secretory or membrane proteins represented by  
XX AAB88317 - AAB89419. Included in the invention are primers  
XX AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the  
XX cDNA sequences of the invention. The invention also includes methods for  
XX the production of antibodies directed against the proteins, and cDNA  
XX sequences, which can be used in vaccines. The polynucleotide sequences  
XX can be used in gene therapy. The polynucleotide sequences and the  
XX proteins they encode may be used in the prevention, treatment and  
XX diagnosis of diseases associated with inappropriate secretory  
XX protein/membrane protein expression. The nucleic acids and complementary  
XX sequences may also be used as DNA probes in diagnostic assays  
XX (e.g. polymerase chain reactions (PCR)) to detect and quantitate the  
XX presence of similar nucleic acid sequences in samples. They may also be  
XX used to study the expression and function of secretory proteins/membrane  
XX polypeptides and their role in metabolism. The polypeptides may be used  
XX as antigens in the production of antibodies against them and in assays to  
XX identify modulators (agonists and antagonists) of expression and  
XX activity. The antibodies and antagonists may also be used as therapeutic  
XX agents to down regulate expression and activity. The antibodies may also  
XX be used as diagnostic agents for detecting the presence of the  
XX polypeptides in samples (e.g. by enzyme linked immunosorbent assay  
XX (ELISA)). Examples of diseases which may be treated include rheumatoid  
XX arthritis and diabetes.  
XX Sequence 2563 BP; 589 A; 753 C; 713 G; 508 T; 0 other;  
XX Query Match 13.8%; Score 156.4; DB 22; Length 2563;  
XX Best Local Similarity 53.6%; Pred. No. 2.6e-25;  
XX Matches 437; Conservative 0; Mismatches 346; Indels 33; Gaps 4;  
Qy 153 CCGCTTGTGCGGACAGGAGGCGCTTCTCCACCGCAACTGAGCGCGCGCGCGGAGCC 212  
Db 255 CCGACCCAGAGTGTCCGCGCGACTTCCGCTGGAGCTGGACCCCGACTGCTCCCGACCC 314  
Qy 213 CGAGCCGCGCGCGCGCTCGCAGCTATCCGAGCTGCGGCTTTCGGGGCGCTGCTGCG 272  
Db 315 GCGCCAGGCGCTCGGGCGCGCGCGCTGCGGAGCTTCTTCGGGGCGCGCTTCTGCG 374  
Qy 273 CCGCGCGCGCTGCTCAAGCGCTGCAAGCGCGCGCTGCGGAGCTTCCGCCAGTCCCGAGCC 332  
Db 375 TCGCGCTGCTGCTGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCTC-----GCT 425  
Qy 333 CAGCGCGAGGTGCTGCGGACTTCCAGCGCGCGCGCGCTCAAGTCTCTGAGTTCG 392  
Db 426 CAGCGAAGAGATGAGTGGAGTTCGCGAAGCGCGCGCGCTTCCAGCTTACCTGCGAGTGC 485  
Qy 393 TTACTTCAGGCAATATCTCCCAAGCGCATGCGCGCTGCTCACACCTTCTACTGAA 452  
Db 486 CTACTTCAAGATCAACAGTTGGAGAGCTGTTGCTGCGAGCACACCTTCTCTGTTGG 545  
Qy 453 GCATCTTGATGACGAATGATGAAGAGGAAATGCGCATATTATAAGAGCGCTGCTGTTG- 511  
Db 546 CAATCTTGAGCACATGGAATGAGCAGACAGCTAGACTATTACCAACCATGCTGGAGT 605  
Qy 512 --CCGAGGACTACATTAAAGCTTGAACCAAGCTATATGAAGCTGTTCTATCCGAGC 569  
Db 606 GAAGAGGCGCGACTTCAAGGATCTTGAGACTTCAACCCCATATGCAAGAAATTGAGCTGG 665  
Qy 570 AGTGGGCGCATACACCGTGTGAGAACTGGAGAAATCCATCCATCAGACATGAGCTGGCGCT 629

Db 666 AGTGGGACTCTACTCAGAGGAAACAGCCACAGAAAGCTGTGCCCACTAGAGGGCGGCT 725  
Qy 630 TCCCGACTTCTTCAAGGCTTTTACGAGTGTCTCGAGCGCTGCGAGGCTTCCAGGAGAT 689  
Db 726 GCAAGAAATACCTTTGTGGCTATGAGAGTGCCTGCGGAGGGCGCTTATGACTA 785  
Qy 690 CAAGGACTTCAA-----GGATTCTACCTTTCCATAGCAGATCATTA 731  
Db 786 CGATGGCTACAACTACCTTGTAGTACAAAGCTGACCTCTTCCAGGCGCATCAGATCATTA 845  
Qy 732 TGTAGAGTCTTGGAAATGCAAAATACAGTGTGAAGAGAACT--CACCCAGTTATAGG 788  
Db 846 CATCCAGTCTCTCACTGTAAGCAGAACTGTGTACGAGGCTTCTTCCACCAAGTCG 905  
Qy 789 AGGTATCCGGTTGAGAAATTTGGTGTACCATATCATTTACTTTCAGTTTGGCTATTA 848  
Db 906 AGAGAAGCGCTTGAAGACTTCTCCCATCGATTATAATATCTGAGTTTGGCTACTA 965  
Qy 849 TAACTTGAACCACTGAGAAATGAGAACTGAGCGCGCTGCGAGTGTGCTGCTTGTATCA 908  
Db 966 TAACTTGGGAATATATACAGGCTGTGAATGTGCCAAGACCTATCTTCTTCTTCTCCC 1025  
Qy 909 GAAATGACAAAGTCTATGCGAGCAAGAACTGGTGTATTA 944  
Db 1026 CAATGACGAGGTGATGACCAAAATTTGGCTATTA 1061  
XX RESULT 15  
XX AAF14589  
XX ID AAF14589 standard; cDNA; 2585 BP.  
XX AC AAF14589;  
XX DT 26-JUN-2001 (first entry)  
XX DE Human cDNA sequence SEQ ID NO:12193.  
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX OS Homo sapiens.  
XX PN EP1074617-A2.  
XX PD 07-FEB-2001.  
XX PF 28-JUL-2000; 2000EP-0116126.  
XX PR 29-JUL-1999; 99JP-0248036.  
XX PR 27-AUG-1999; 99JP-0300253.  
XX PR 11-JAN-2000; 2000JP-0118776.  
XX PR 02-MAY-2000; 2000JP-0183767.  
XX PR 09-JUN-2000; 2000JP-0241899.  
XX PA (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX Claim 8; SEQ ID 12193; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
XX to the complementary strand of a polynucleotide which comprises one of  
XX the 5602 nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2585 BP; 590 A; 759 C; 723 G; 513 T; 0 other;

Query Match 13.8%; Score 156.4; DB 22; Length 2585;  
Best Local Similarity 53.6%; Pred. No. 2.6e-25;  
Matches 437; Conservative 0; Mismatches 346; Indels 33; Gaps 4;

Qy	153	CCGCTTCTGCGCACAGCAGGCGCTTCTGCCACCGCAACTGCAGCGCGCGCGCGCAGCC	212
Db	277	CCGCACCCAGTGTGCCCGCGACTTCCCGTGGAGCTGGACCCCGACTGGTCCCCAGGCC	336
Qy	213	CGAGCCCGCGCGCGCGCTCCCGAGCTATCCCGAGCTGCGGCTCTTCGGGGCGCTGCTGG	272
Db	337	GGCCAGGCTCGGGCGCGCGCGCTCGCGACCTGAGCTTCTTCGGGGCGCTTCTGCG	396
Qy	273	CCGGCGCGCACTGCTCAAGCGCTGCAAGCAGGGGCTGCCAGCGCTTCGCCAGTCCCGAGCC	332
Db	397	TGCGCTGCTGCTGCTGCGCGCTGCTGCGCGCGCGCGCGCGCACTC-----GCT	447
Qy	333	CAGCCCGAGGTGTCGCGGACTTCCAGCGCGCGAGCCCTACAAGTTCCTGCACTGCG	392
Db	448	CAGCGAAGATGAGCTGGAGTTCCGCAAGCGGAGCGCCCTACAACTACTGCGAGTGG	507
Qy	393	TTACTTCAAGCAATATCTCCCAAGCATCGCGCTGCTCAGACCTTCTACTGAA	452
Db	508	CTACTTCAGATCAACAAGTTGGAGAAAGCTGTTGTCGAGCACACACCTTCTTCTGGG	567
Qy	453	GCATCTTGATGAGAAATGATGAAGAGGAACATGGCATATTATAAGAGCTGCTGGTG-	511
Db	568	CAATCTTGAGCACATGGAATGAGCAGAACCTAGACTATTACCAACCATGCTCGAGT	627
Qy	512	--CGAGGACTACATTAAGACCTGGNAACCAAGTCATATGAAGCGCTGTTCAATCCGAGC	569
Db	628	GAAGGAGCGCGACTTCAAGGATCTTGAGACTCAACCCCATATGCAAGAATTTGACTGG	687
Qy	570	AGTGGGGGCATACAAACGGTGAGAACTGGAGAACATCCATCACAGACATGGAGTGGCCCT	629
Db	688	AGTGGACTCTACTCAGAGGAAACAGCCACAGGAAGCTGTGCCCACTAGAGCGCGCT	747
Qy	630	TCCGAGCTTCTTAAAGCCTTTTACAGTGTCTCGAGGCTGCGAGGGTCCAGGGAGAT	689
Db	748	GCAAGATATCTTTGTGCGCTATGAGGAGTCCCGTGGCCCTCTGCAAGGGCCCTATGACTA	807
Qy	690	CAAGGACTTCAA-----GGATTCTACCTTTCCATAGCAGATCATT	731
Db	808	CGATGGCTAACACTCCTTGAGTACACGCTGACCTCTTCCAGGCCATCACAGATCATT	867
Qy	732	TGTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAGAACCT---CACCCAGTTATAGG	788
Db	868	CATCCAGGTCTCAACTGTAAAGCAGAACTGTGTACGGAGCTTGCTTCCCAACCCAGTCG	927
Qy	789	AGGCTATCCGGTTGAGNAATTTGGGCTACCATGATCATTAATCTGAGCTTGCCTATTA	848
Db	928	AGAGAGCCCTTTGAGAGCTTCTCCCATCGCATTAATTAATCTGAGCTTGCCTACTA	987
Qy	849	TAAAGTTGAACGAGCTGAAGAAATGACGCCCGCTGTCAGTCAGCTATCTGCTTTGATCA	908

Db	988	TAACTTGGGAATTATACACAGGCTTTGAATGTGCCAAGACCTATCTTCTTCTCC	1047
Qy	909	GAATGACAAGGTCAATGCAGCAAGCTGCTGCTATTA	944
Db	1048	CAATGACGAGGTGATGAACCAAAATTTGGCCTATTA	1083

Search completed: January 30, 2004, 08:32:19  
Job time : 297.825 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 08:13:15 ; Search time 406.168 Seconds  
(without alignments)  
10794.564 Million cell updates/sec

Title: US-09-729-674-1\_COPY\_63\_1265

Perfect score: 1203

Sequence: 1 atggagcggggcgccggggg.....tgggaactgggagaccagc 1203

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1202.2	99.9	3871	9	US-09-729-674-1
2	919	76.4	1810	12	Sequence 1, Appli
3	467	38.8	2443	13	Sequence 2058, Ap
4	465.8	38.7	2347	10	Sequence 724, App
5	465.8	38.7	2347	10	Sequence 113, App
6	195.6	16.3	398	12	Sequence 64, Appl
7	154.8	12.9	2800	14	Sequence 8221, Ap
8	153.8	12.8	892	13	Sequence 1, Appli
9	153.8	12.8	892	13	Sequence 161474,
10	153.8	12.8	892	13	Sequence 161475,
11	153.8	12.8	892	13	Sequence 161476,
12	153.8	12.8	892	14	Sequence 161474,
13	153.8	12.8	892	14	Sequence 161475,
14	153.8	12.8	892	14	Sequence 161476,
15	147.4	12.3	2322	14	Sequence 7, Appli

16	147.4	12.3	2416	14	US-10-045-815-5	Sequence 5, Appli
17	145.4	12.1	564	13	US-10-027-632-114858	Sequence 114858,
18	145.4	12.1	564	14	US-10-027-632-114858	Sequence 114858,
19	131.6	10.9	3396	13	US-10-094-749-921	Sequence 921, App
20	126	10.5	618	10	US-09-879-536-215	Sequence 215, App
21	109.2	9.1	2665	14	US-10-071-785-86	Sequence 86, Appl
22	107.2	8.9	926	12	US-10-242-535A-9365	Sequence 9365, Ap
23	103.4	8.6	150	12	US-10-242-535A-1775	Sequence 1775, Ap
24	97.8	8.1	2753	10	US-09-728-952-28	Sequence 28, Appl
25	97.8	8.1	2753	10	US-09-728-952-51	Sequence 51, Appl
26	87	7.2	349	9	US-09-925-302-395	Sequence 395, App
27	73	6.1	247	12	US-10-242-535A-4919	Sequence 4919, Ap
28	69.6	5.8	181	13	US-10-023-386-16058	Sequence 16058, A
29	69.6	5.8	181	13	US-10-023-386-16058	Sequence 2388, Ap
30	68.2	5.7	3897	13	US-10-023-386-2358	Sequence 7, Appli
31	68.2	5.7	3897	14	US-10-021-657-7	Sequence 1, Appli
32	60	5.0	1906	13	US-10-412-000-1	Sequence 1, Appli
33	60	5.0	1906	14	US-10-021-657-1	Sequence 22, Appl
34	59.8	5.0	2298	10	US-09-476-242-22	Sequence 23, Appl
35	59.8	5.0	2298	10	US-09-476-242-23	Sequence 24, Appl
36	59.8	5.0	2298	10	US-09-476-242-24	Sequence 21, Appl
37	59.8	5.0	2310	10	US-09-476-242-21	Sequence 4, Appli
38	59.8	5.0	2310	10	US-09-476-242-4	Sequence 8, Appli
39	59.8	5.0	2316	10	US-09-476-242-5	Sequence 5, Appli
40	59.8	5.0	2322	10	US-09-476-242-8	Sequence 18, Appl
41	59.8	5.0	2322	10	US-09-476-242-18	Sequence 19, Appl
42	59.8	5.0	2322	10	US-09-476-242-19	Sequence 20, Appl
43	59.8	5.0	2322	10	US-09-476-242-20	Sequence 6, Appli
44	59.8	5.0	2322	10	US-09-476-242-20	
45	59.8	5.0	2328	10	US-09-476-242-6	

## ALIGNMENTS

RESULT 1  
US-09-729-674-1  
; Sequence 1, Application US/09729674  
; Patent No. US30010039335A1  
; GENERAL INFORMATION:  
; APPLICANT: McCoy, John M.  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: Lavallie, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steininger II, Robert J.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fectel, Kim  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: 6055-64X  
; CURRENT APPLICATION NUMBER: US/09/729,674  
; CURRENT FILING DATE: 2000-12-04  
; PRIOR FILING DATE: 2000-03-30  
; NUMBER OF SEQ ID NOS: 283  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 3871  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-729-674-1

Query Match 99.9%; Score 1202.2; DB 9; Length 3871;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGAGCGGGCGCGGGGGCGGGCGGTGTAGCGCTGTGTGCGTGGCGCGCG



QY 601 CTGGAACCAAGTCATATGAAGCCCTGTTTCATCCGAGCAGTCCGGGCATACAAACGGTGAG 660  
Db 624 CTGGAACCAAGTCATATGAAGCCCTGTTTCATCCGAGCAGTCCGGGCATACAAACGGTGAG 683  
QY 661 AACTGGAGACATCCATCAGACATGAGCTGGCCCTTCCCGACTTCTTCAAAGCCCTTT 720  
Db 684 AACTGGAGACATCCATCAGACATGAGCTGGCCCTTCCCGACTTCTTCAAAGCCCTTT 743  
QY 721 TAGCAGTGTCTCGCAGCCCTGCCAGGCTTCCAGGAGATCAAGGACTTCAAGGATTTCTAC 780  
Db 744 TAGCAGTGTCTCGCAGCCCTGCCAGGCTTCCAGGAGATCAAGGACTTCAAGGATTTCTAC 803  
QY 781 CTTTCCATAGCAGATCATTTATGTAGAACTTCTGGAATGCAAAATACAGTGTGAAGAGAAC 840  
Db 804 CTTTCCATAGCAG----- 816  
QY 841 CTCACCCAGTTATAGGAGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTTAC 900  
Db 817 ----- 816  
QY 901 TTGCAGTTTGCCTATTATAAGTTGAACGACCTGAAGAAATGACGCCCTGTGCAGTCAGC 960  
Db 817 -----TGAACGACCTGAAGAAATGACGCCCTGTGCAGTCAGC 854  
QY 961 TATCTGCTCTTTGATCAGATGACAAAGTGTGATGACGAGAACTTGTGTATTACAGTAC 1020  
Db 855 TATCTGCTCTTTGATCAGATGACAAAGTGTGATGACGAGAACTTGTGTATTACAGTAC 914  
QY 1021 CACAGGACACTTTGGGGCTCTCGGATGAGCACTTCCAGCCAGACCTGAAAGCAGTTTCAG 1080  
Db 915 CACAGGACACTTTGGGGCTCTCAGATGAGCACTTCCAGCCAGACCTGAAAGCAGTTTCAG 974  
QY 1081 TTCTTTAATGTGACACACTCAGAAAGAGCTGTATGACTTTGCTTAAGAAAAATATAATG 1140  
Db 975 TTCTTTAATGTGACACACTCAGAAAGAGCTGTATGACTTTGCTTAAGAAAAATATAATG 1034  
QY 1141 GATGATGATGAGGAGAACTTGTGAAATATGGAATGATCCTCTTGAACTGAGGAGACC 1200  
Db 1035 GATGATGATGAGGAGAACTTGTGAAATATGGAATGATCCTCTTGAACTGAGGAGACC 1094  
QY 1201 AGC 1203  
Db 1095 AGC 1097

## RESULT 3

US-10-094-749-724  
; Sequence 724, Application US/10094749  
; Publication No. US20030219741A1

; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO

; APPLICANT: SUGIYAMA, TOMOYASU

; APPLICANT: OTSUKI, TETSUJI

; APPLICANT: WAKAMATSU, AI

; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO

; APPLICANT: YAMAMOTO, JUN-ICHI

; APPLICANT: ISONO, YUUKO

; APPLICANT: HIO, YURI

; APPLICANT: OTSUKA, KAORU

; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, RYOTARO

; APPLICANT: TAMECHIKA, ICHIRO

; APPLICANT: SEKI, NAOHICO

; APPLICANT: YOSHIKAWA, TSUTOMU

; APPLICANT: OTSUKA, MOTOKYUKI

; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA

; FILE REFERENCE: 084335/0160

; CURRENT APPLICATION NUMBER: US/10/094,749

; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 60/350,435

; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381

; PRIOR FILING DATE: 2001-09-14

; NUMBER OF SEQ ID NOS: 3381

; SOFTWARE: PatentIn ver. 2.1

; SEQ ID NO 724

; LENGTH: 2443

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-094-749-724

Query Match

Best Local Similarity 38.8%; Score 467; DB 13; Length 2443;

Matches 755; Conservative 2; Mismatches 402; Indels 18; Gaps 3;

QY 8 CCGGGCCCGGGGGCGCGCGCTGCTAGCGTGTGTGCGTGGCTGCGCGCTGCGCG 67  
Db 451 CCGGGCCCGGGGGCGCGCGCATGGCTCGGGTGGCGGGGCTGCTGTGCTGCTGGCA 510  
QY 68 CCGGGCCCGGGGGCGCGCGCTGCTAGCGTGTGCGAGCTTCCACGGGACGAGCTGATGC 127  
Db 511 GCGCGGGGGCGCAGTACGAGAGTACAGCTTCCGGGGCTTCCCGCCGAGGACCTGATGC 570  
QY 128 CGCTCGAGTCCGCTTACCGGACGCGCTGGACAAGTACAGCGCGGAGCACTGGGCCGAGA 187  
Db 571 CGTGGCCCGGGGGCTACGGGCACTGCTGGAGCAGTACGAGGAGAGAGCTGGCGCGAGA 630  
QY 188 GCSTTGGCTACCTTGGAGATCAGCGCTGCGGCTGACCGCTTGTCTGCGCGACGAGCGCCT 247  
Db 631 CCGCGCGCTACCTTGGAGGCGCGCTGCGGCTGACCGCTTGTCTGCGCGACGAGGCGCT 690  
QY 248 TCTGCCACCGCACTGAGCG-----CGCGCGCGAGCGCGCGCGCGCGCGCGCC 298  
Db 691 TCTGCCACCGCACTGAGCC 750  
QY 299 TCGCAGCTATC-----CGAGCTGCGCTCTTGGGGGCGCTGCTGCGCGCGCGCGCACT 352  
Db 751 GCGCAGCAGTGGGCTGCGAGCTGCGGCTCTTGGCGCGCTTCTGAGCGGAGCGCGCT 810  
QY 353 GCGTAAAGCGCTCAAGCAGGCGCTGCGAGCTTCCCGAGTCCCGAGTCCCGAGCGCGGAGG 412  
Db 811 GCGTGGCGCTGCAAGCGGAGCGCTGCGCGCTTCCAGGTGCGCTACCGCGCGCGCGCAGC 870  
QY 413 TGCTGGCGACTTCCAGCGCGCGCGCGCTTCCAGGTGCGCTTCTGAAAGCATCTCTGATG 472  
Db 871 TGCTGGTGTACTTCCAGAGCGCGCTGCGCTTCCAGGTGCGCTTCTGAAAGCATCTCTGATG 930  
QY 473 CAATAATCTCCCGCAAGCGCATCGCGCTGCTCACACCTTCTTCTGAAAGCATCTCTGATG 532  
Db 931 CTAAACCGCTGGAGAGGCGGTGGCGCGCGCTTACACCTTCTTCCAGAGGAAACCGAAGC 990  
QY 533 ACGAAATGATGAAGAGGAAACATGGCATATTATAA---GAGCGTGCCTGTGCGGAGGACT 589  
Db 991 ACGAGCTGACCGCAAGATATCTCACTACTATCAGGGGATGCTGGACGCTGCGCGAGGACT 1050  
QY 590 ACATTAAGACCTGGAAACCAAGTCATATAAGACCTGTTTCATCCGAGCAGTGGCGGCACT 649  
Db 1051 CCCTCAGGACCTAGAGGCCAGCCCTACAGCGCGTGTCTTCCGGGTGTGAAGCTCT 1110  
QY 650 ACAACGGTGAAGACTGGAGAACATCCATCAGACATGAGCTGGCGCTTCCCGACTTCT 709  
Db 1111 ACAACAGCGGGGATTTCCGAGCAGCAGCAGGAGACATGAGCGGGGCTTGTTCAGAGTACC 1170  
QY 710 TCAAGCCCTTTACGAGTGTCTCGAGCGCTGGAGGTTCCAGGAGATCAAGGACTTCA 769  
Db 1171 TGGCAGTCTTTGCCCGGCTGGCGGCTGTGAGGGGCGCCATGAGCAGGTGGAGTCTCA 1230  
QY 770 AGGATTTCTACCTTTCCATAGCAGATCATATTGTAGAAGTCTTGGAAATCAAAATACAGT 829  
Db 1231 AGGATTTCTACCGCGCCATAGCAGATCTCTTTGAGAGTCCCTGCGAGTCAAGGTGACT 1290  
QY 830 GTGAAGAGAACCTCACCCAGTTATAGAGGCTATCCCGTTGAGAAATTTGTGGTATCCA 889









Db 627 GAAGAGCGGACCTTCAAGAGTCTTGAGACTCAACCCCATATGACGAATTTGCACTGGG 696  
QY 639 AGTGGGCGCATACACCGGTGAGAACTGAGAACTCCATCACAGACATGAGCTGGCCCT 698  
Db 687 AGTGGGACTCTACTCAGAGGACACCCACAGGAGCTGTGCCACCTAGAGCGGGCT 746  
QY 699 TCCGAGCTTCTTCAAGCTTTTACGAGTGTCTGCGAGCTGCGAGGTTCCAGGGAGAT 758  
Db 747 GCAAGAACTACTTTGTGGCCTATGAGGAGTGGCGCTCTGCGAAGGGCCCTATGACTA 806  
QY 759 CAAGGACTTCAA-----CGATTTCTACCTTTCCATAGCAGATCATTA 800  
Db 807 CGATGGCTACAACTACCTTTGAGTACACCGCTGACTCTTCCAGGCGCATCAGATCATTA 866  
QY 801 TGTAGAATTTCTGGAATGCAAAATACAGTGTGAAGAAAGCT---CACCCAGTTATAGG 857  
Db 867 CATCCAGTCTCCTAACTGTAAAGAGAACTGTGTACGAGGCTTGTCTTCCCAACCAAGTCG 926  
QY 858 AGGCTATCCGTTTGCAGAAATTTGGGTACCATGTATCTACTTCTGAGTTTCCCTATTA 917  
Db 927 AGAAGAGCCCTTTGAGAGCTTCTCCCATGCGATGCAATTAATATCTGAGTTTGCCTACTA 986  
QY 918 TAAGTTGAACGACCTGAGAAATGAGCCCTGTGCGAGTCACTATCTGCTCTTTGATCA 977  
Db 987 TAACATTTGGGAATTTATACACAGCTGTGTGATGTGCGCAAGACCTATCTTCTTCTTCCC 1046  
QY 978 GAATGACAGGCTCAGCAGAGACCTGGTGATTA 1013  
Db 1047 CAATGACAGGTGATGAACCAAAATTTGGGCTATTA 1082

## RESULT 9

US-10-027-632-161474/c

; Sequence 161474, Application US/10027632

; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027.632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 161474

; LENGTH: 892

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-161474

Query Match 12.8%; Score 153.8; DB 13; Length 892;  
Best Local Similarity 97.5%; Pred. No. 1.9e-34;  
Matches 155; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 463 TACTTCAAGGCAATAATCTCCCAAGCCATCGCCGCTGCTCACACCTTTCTACTGAAG 522  
Db 228 TCCTTTATAGGCAATAATCTCCCAAGCCATCGCCGCTGCTCACACCTTTCTACTGAAG 169  
QY 523 CATCTGTATGACGAATGATGAAGAGGAACATGGCATATTATAGAGGCTGCTGCTGCC 582

Db 168 CATCTGTATGACGAATGATGAAGAGGAACATGGCATATTATAGAGGCTGCTGCTGCC 109  
QY 583 GAGGACTACATTAAAGACCTGGAACCAAGTCATATGAA 621  
Db 108 GAGGACTACATTAAAGACCTGGAACCAAGTCATATGAA 70

## RESULT 10

US-10-027-632-161475/c

; Sequence 161475, Application US/10027632

; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027.632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 161475

; LENGTH: 892

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-161475

Query Match 12.8%; Score 153.8; DB 13; Length 892;  
Best Local Similarity 97.5%; Pred. No. 1.9e-34;  
Matches 155; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 463 TACTTCAAGGCAATAATCTCCCAAGCCATCGCCGCTGCTCACACCTTTCTACTGAAG 522  
Db 228 TCCTTTATAGGCAATAATCTCCCAAGCCATCGCCGCTGCTCACACCTTTCTACTGAAG 169  
QY 523 CATCTGTATGACGAATGATGAAGAGGAACATGGCATATTATAGAGGCTGCTGCTGCC 582  
Db 168 CATCTGTATGACGAATGATGAAGAGGAACATGGCATATTATAGAGGCTGCTGCTGCC 109  
QY 583 GAGGACTACATTAAAGACCTGGAACCAAGTCATATGAA 621  
Db 108 GAGGACTACATTAAAGACCTGGAACCAAGTCATATGAA 70

## RESULT 11

US-10-027-632-161476/c

; Sequence 161476, Application US/10027632

; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027.632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 161476  
;; LENGTH: 892  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-161476

Query Match 12.8%; Score 153.8; DB 13; Length 892;  
Best Local Similarity 97.5%; Pred. No. 1.9e-34;  
Matches 155; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 463 TACTTCAAGGCAAAATAATCTCCCAAGCCATCGCGCTGTCTCACACCTTTCTACTGAAG 522  
Db 228 TCCTTTTAGGCAAAATAATCTCCCAAGCCATCGCGCTGTCTCACACCTTTCTACTGAAG 169

QY 523 CATCTGTATGACGAATGATGAAGGACATGGCATATTATAGAGCCTCGCTGTGCCC 582  
Db 168 CATCTGTATGAYGAATGATGAAGGACATGGCATATTATAGAGCCTCGCTGTGCCC 109

QY 583 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 621  
Db 108 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 70

RESULT 12  
US-10-027-632-161476/c  
;; Sequence 161474, Application US/10027632  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 161474  
;; LENGTH: 892  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-161474

Query Match 12.8%; Score 153.8; DB 14; Length 892;  
Best Local Similarity 97.5%; Pred. No. 1.9e-34;  
Matches 155; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 463 TACTTCAAGGCAAAATAATCTCCCAAGCCATCGCGCTGTCTCACACCTTTCTACTGAAG 522

Db 228 TCCTTTTAGGCAAAATAATCTCCCAAGCCATCGCGCTGTCTCACACCTTTCTACTGAAG 169

QY 523 CATCTGTATGACGAATGATGAAGGACATGGCATATTATAGAGCCTCGCTGTGCCC 582  
Db 168 CATCTGTATGAYGAATGATGAAGGACATGGCATATTATAGAGCCTCGCTGTGCCC 109

QY 583 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 621  
Db 108 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 70

RESULT 13  
US-10-027-632-161475/c  
;; Sequence 161475, Application US/10027632  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 161475  
;; LENGTH: 892  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-161475

Query Match 12.8%; Score 153.8; DB 14; Length 892;  
Best Local Similarity 97.5%; Pred. No. 1.9e-34;  
Matches 155; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 463 TACTTCAAGGCAAAATAATCTCCCAAGCCATCGCGCTGTCTCACACCTTTCTACTGAAG 522  
Db 228 TCCTTTTAGGCAAAATAATCTCCCAAGCCATCGCGCTGTCTCACACCTTTCTACTGAAG 169

QY 523 CATCTGTATGACGAATGATGAAGGACATGGCATATTATAGAGCCTCGCTGTGCCC 582  
Db 168 CATCTGTATGAYGAATGATGAAGGACATGGCATATTATAGAGCCTCGCTGTGCCC 109

QY 583 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 621  
Db 108 GAGGACTACATTAAAGACCTGGAAACCAAGTCATATGAA 70

RESULT 14  
US-10-027-632-161476/c  
;; Sequence 161476, Application US/10027632  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676



Result No.	Query Match	Score	*	Length	DB	ID	Description
C 1	126	10.5	618	3	US-09-328-111-215	Sequence 215, Appl	
C 2	64	5.3	846	4	US-09-253-991A-1177	Sequence 1177, Ap	
C 3	64	5.3	1083	4	US-09-253-991A-1387	Sequence 1387, Ap	
C 4	64	5.3	1089	4	US-09-253-991A-1221	Sequence 1221, Ap	
C 5	60.4	5.0	30001	1	US-08-125-468-1	Sequence 1, Appl	
C 6	60.4	5.0	30001	2	US-08-474-933-1	Sequence 1, Appl	
C 7	59	4.9	2481	1	US-08-324-243-35	Sequence 35, Appl	
C 8	59	4.9	2481	1	US-08-533-390-35	Sequence 35, Appl	
C 9	59	4.9	2481	3	US-08-717-294-35	Sequence 35, Appl	
C 10	59	4.9	2481	5	PCF-U895-11511-35	Sequence 35, Appl	
C 11	59	4.9	2571	4	US-09-553-950-4	Sequence 4, Appl	
C 12	55.4	4.6	1140	3	US-09-105-537-15	Sequence 15, Appl	
C 13	55.4	4.6	13613	3	US-09-105-537-3	Sequence 3, Appl	
C 14	54.4	4.5	3292	3	US-09-320-878-22	Sequence 22, Appl	
C 15	54.4	4.5	3292	4	US-09-141-908-15	Sequence 15, Appl	
C 16	54.4	4.5	3292	4	US-09-657-440-22	Sequence 22, Appl	
C 17	52.6	4.4	1686	4	US-09-253-991A-6509	Sequence 6509, Ap	
C 18	52.6	4.4	1740	4	US-09-253-991A-6553	Sequence 6553, Ap	
C 19	52.6	4.4	1489	4	US-09-253-991A-6558	Sequence 6558, Ap	
C 20	52.6	4.4	2469	4	US-09-253-991A-6519	Sequence 6519, Ap	
C 21	52	4.3	1155	4	US-09-253-991A-15330	Sequence 15330, A	
C 22	52	4.3	3624	1	US-07-951-715A-6	Sequence 6, Appl	
C 23	52	4.3	3624	2	US-08-459-448A-6	Sequence 6, Appl	
C 24	52	4.3	3624	3	US-08-459-595A-6	Sequence 6, Appl	
C 25	52	4.3	3624	3	US-08-459-504B-6	Sequence 6, Appl	
C 26	52	4.3	3624	3	US-08-459-444-6	Sequence 6, Appl	
C 27	52	4.3	3624	3	US-09-053-549-7	Sequence 7, Appl	





TELEFAX: (201)831-3305  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-125-468-1

Query Match 5.0%; Score 60.4; DB 1; Length 30001;  
Best Local Similarity 47.9%; Pred. No. 8.6e-05;  
Matches 203; Conservative 1; Mismatches 217; Indels 3; Gaps 1;

QY 39 GCTGCTGTGGTGGCTGTGGCGCTGGCGCCGGCGCGGCCCAATACGACAGCTACAGTT 98  
Db 6270 GTTGCTGTGGCGAGCGCACCGCCGAATCTGGCGTCCCCGTGGAATGGGGCTCGAACTGC 6211  
QY 99 CCGCAGCTTCCCACGGGACGAGCTGATGCCGCTCGAGTCGGCTACC GGACGGCTGGA 158  
Db 6210 CGCGCTGCAGCCCGGSCCGACTCCGTGCTCATCCCTCCGCCACGCCGACGGCGCAC 6151  
QY 159 CAAGTACAGCGCGAGACATGTGGCCGAGAGSTKGGCTACTCTGGAGATCAGCTTCGGGCT 218  
Db 6150 CGAGGAGACCCGGTGGCTGGCTGGTGGCGCCGACGCGCGCGCAGCACCGTGGCGCAA 6091  
QY 219 GCACCGCTTGTGCGCGACAGGAGGCGCTTGTGCCACCGCAACTGCAGCGCGCGCCCGCA 278  
Db 6090 GCAGCTGGGATGGCGCTGATCGGCGAATCCACCGAGTCTGGCTCAACGGCGAGCTCAC 6031  
QY 279 GCCCGAGCCCGCGCGGCGCTCGCCAGCTATCCGAGCTGGCGCTCTTCGGGGGCGCTGCT 338  
Db 6030 CCTGACGCGGACCTCCGCGCGACAGCAACACCTGTGCACACCGAGCCCGCACCT 5971  
QY 339 GCGCGCGCGCACTGCTCAAGCGTGAAGAGGGCTGCAGAGGGCTTCGCCCAAGTCCA 398  
Db 5970 GTGTGCTGTGCTGCTTCCCGGAGCGCGGCAAGTGGCGGTGTGTGACACGGTGGACACCGA 5911  
QY 399 GCCCA3CGCGGAGGTGTGCGGCGACTTCCAGCGCGCGAGCGCTTACAAGTTCTTCGCAGTT 458  
Db 5910 ---CGGCGCGAGGACCGCGAGCTGGTCCGCGCGCGCTGCCCGACAGAATCGGGCGGG 5854  
QY 459 CGCT 462  
Db 5853 CGCT 5850

RESULT 6  
US-08-474-933-1/c  
Sequence 1, Application US/08474933  
Patent No. 5866410  
GENERAL INFORMATION:  
APPLICANT: Ryan, Michael J.  
APPLICANT: Lotvin, Jason A.  
APPLICANT: Strathy, Nancy E.  
APPLICANT: Fantini, Susan E.  
TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
TITLE OF INVENTION: chlorotetracycline and tetracycline Formation and cosmid  
TITLE OF INVENTION: useful therein  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESS: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07470  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:



RESULT 9  
US-08-7117-294-35  
; Sequence 35, Application US/08717294  
; Patent No. 6114148  
; GENERAL INFORMATION:  
; APPLICANT: SEED, BRIAN  
; APPLICANT: HAAS, JURGEN  
; TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF  
; TITLE OF INVENTION: PROTEINS  
; NUMBER OF SEQUENCES: 110  
; CORRESPONDENCE ADDRESS:

Query Match	4.9%; Score 59; DB 3; Length 2481;
Best Local Similarity	48.2%; Pred. No. 7.8e-05;
Matches 198; Conservative	0; Mismatches 210; Indels 3; Gaps 1;
QY	31 CTGCTAGCGGTGCTGTGGCGCTCGCGCGCTGC GGCGCGCGCCCAATACGAACGC 90
Db	1990 CTGTTGGGCTTCGGATCGTGTTCGCCGCTGCTAGCATCGTAGAACCGCTGCGCCACGGGC 2049
QY	91 TACAGCTTCGCGAGCTTCCCACGGGACGAGCTGATGCCGCTCGAGTCGGCTTACCGGCAC 150
Db	2050 TACAGCCCCGTAGCCTCCAGACCCGGCCCCCGTGCGCGGGGCCGACCCGCCCGCGAG 2109
QY	151 GCGCTGGACAAGTACAGCGCGCGAGCACTGGGCGGAGAGCTTGGCTACTCTGGAGATCAGC 210
Db	2110 GGCATCGAGGAGGAGGGCGCGAGCGCGACCGCGACACAGCGCGGAGGCTCGTGTGACGGC 2169
QY	211 CTGGCGCTGCACCGCTTGCTGGCGGACACGAGGCGCTTCTGCCACCGCAACTGCACGCGC 270
Db	2170 TTCTTGGCGATCATCTGGGTGCAGCTCTCCGAGCCTGTTCTGTTTACGCTTACCAACCGC 2229
QY	271 GCGCGCAGCCGAGCCCGCGCGCGCTTCGCCAGGTATCCGAGCTGCGCTCTTTCGGG 330
Db	2230 GACCTGCTGCT---GATCCGCCCGCCGATCGTGGAACTCTTAGCGCGCGCGCTGGGAG 2286
QY	331 GGCCTGCTCGCGCGCGCGACCTGCTCAAGCGCTGCAAGCAGGGGCGCTGCCACGCCCTCCGC 390
Db	2287 GTGCTGAAGTACTGTGTGNACCTCTCTCAGTATTGGAGCCAGAGCTGAAGTCCAGCGCC 2346
QY	391 CAGTCCGAGCCCGCGGAGGTGTGGCGGACTTCCAGCGCGCGGAGGCC 441
Db	2347 GTGAGCGCTGTGAACGCCACCGCCCATCGCTGTGGCGGAGGACCGACCGC 2397

RESULT 10  
PCT-US95-11511-35  
; Sequence 35, Application PC/TUS9511511  
; GENERAL INFORMATION:  
; APPLICANT: SEED, BRIAN



QY 51 GGCTGCGGCTGCGCGCGCGCGCGCCAAATACGACGCTACAGCTTCGGAGCTTCCC 110  
Db |||||  
QY 726 GGCGCGCTGCTGCGGATCCGGCTCCGCCACCTCGGACAGCTGGAACGGCGGAGTCCGC 785  
Db |||||  
QY 111 ACGGACGAGCTGATCCGCTCGAGTCGGCCTACCGGACCGGCTGGCAAGTACAGCGG 170  
Db |||||  
QY 786 GCTGCGCGGAGTACCTCTCCGGGCTCCCGGACTGCCCGCATCGGCCTGCCGTGAC 845  
Db |||||  
QY 171 CGAGCACTGGGCGGAGAGSTYGGCTACTCTGGAGATCAGCTTCGGGCTGCACCGTTGCT 230  
Db |||||  
QY 846 CGCGCCCGACACGACCGCGTCTGGCACCTCTTACCGGTGCGCACCGAGCGCGCGGCGA 905  
Db |||||  
QY 231 GCGGACACGAGGCGCTTCTGCCACCGCACTGCGAGCGCGCGCGCGCGAGCCCGCGC 290  
Db |||||  
QY 906 GCTGCGCACCCCTCGAGCGCGCGGCGATCGACACCTTCAGCACTACCGGTACCCGT 965  
Db |||||  
QY 291 CGCGGCGCTGCGAGTATCCGAGCTGCGCTCTTCGGGGCTGCTGCGCGCGCGCA 350  
Db |||||  
QY 966 GCACCTCTCGCCCGCTACCGCGGCGAGCGCACCGCGGAAAGCTCGCTCCCGCGGCGA 1025  
Db |||||  
QY 351 CTGCTCAAGCGCTGCAAGCAGGCGCTGCGAGCGCTTCGCCAGTCCCGAGCCGCGCGA 410  
Db |||||  
QY 1026 GAGCTTCGCGGCGAGTCTCAGCTTCGCGATCGCGCGCACCTGAGCGCGCGCAGGC 1085  
Db |||||  
QY 411 GGTGCTGCGGACTTCCAGCGCGCGA 437  
Db |||||  
QY 1086 GCTGCGGCTGATCGACGCGCTGCGCGA 1112  
Db |||||

## RESULT 13

US-09-105-537-3/c

; Sequence 3, Application US/09105537A

; Patent No. 6265202

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/105,537A

; CURRENT FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 3

; LENGTH: 13613

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-105-537-3

Query Match 4.6%; Score 55.4; DB 3; Length 13613;  
Best Local Similarity 46.3%; Pred. No. 0.001;  
Matches 179; Conservative 1; Mismatches 207; Indels 0; Gaps 0;

QY 51 GGCTGCGGCTGCGCGCGCGCGCGCCAAATACGACGCTACAGCTTCGGAGCTTCCC 110  
Db |||||  
QY 7244 GGCGCGCTGCTGCGGATCCGGCTCCGCCACCTCGGACAGCTGGAACGGCGGAGTCCGC 7185  
Db |||||  
QY 111 ACGGACGAGTATGCGGCTCGAGTCGGCCTACCGGACCGGCTGGCAAGTACAGCGG 170  
Db |||||  
QY 7184 GCTGCGCGGAGTACCTCTCCGGCTCGCGGACTGCCCGCATCGGCCTGCCGTGAC 7125  
Db |||||  
QY 171 CGAGCACTGGGCGGAGAGCTGAGTACCTGAGATCAGCTTCGGGCTGACCGCTTGTCT 230  
Db |||||  
QY 7124 CGCGCCCGAACCGGCTGCTGCGACCTTTTACCGTGGCACCGAGGCGCGGACGA 7065  
Db |||||  
QY 231 GCGGACACGAGGCGCTTCTGCCACCGCAATGCGAGCGCGCGCGCGAGCCCGAGCCCGC 290  
Db |||||  
QY 7064 GCTGCGGAGCACCTCGAGCGCGCGGATCGACACCTTCAGCACTACCGGTACCCGT 7005  
Db |||||  
QY 291 CGCGGCTGCGGAGTATCCGAGCTGCGCTCTTCGGGGGCTGCTGCGCGCGCGCA 350  
Db |||||  
QY 7004 GCACCTCTCGCCCGCTTACCGGCGGAGGACCGCGGAGGCTCGCTCCCGGCGCGA 6945  
Db |||||

QY 351 CTGCTCAAGCGCTGCAAGCAGGCGCTTCAGGCTTCGGCAGTCCCGAGCCGCGCA 410  
Db |||||  
QY 6944 GAGCTTCGCGCGGCGAGGTCTCAGCTTCGCGATCGGCCCGCACTGGAGGCGCGGAGGC 6885  
Db |||||  
QY 411 GGTGCTGCGGAGCTTCAGGCGCGCGA 437  
Db |||||  
QY 6884 GCTGCGGCTGATCGACGCGCGTGGCGGA 6858  
Db |||||

## RESULT 14

US-09-320-878-22

; Sequence 22, Application US/09320878A

; Patent No. 6117659

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.

; APPLICANT: BETLACH, Mary C.

; APPLICANT: McDANIEL, Robert

; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

; FILE REFERENCE: 300622002120

; CURRENT APPLICATION NUMBER: US/09/320,878A

; CURRENT FILING DATE: 1999-05-27

; EARLIER APPLICATION NUMBER: CIP OF 09/141,908

; EARLIER FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538

; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247

; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: 60/119,139

; EARLIER FILING DATE: 1999-02-08

; EARLIER APPLICATION NUMBER: 60/100,880

; EARLIER FILING DATE: 1998-09-22

; EARLIER APPLICATION NUMBER: 60/087,080

; EARLIER FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 22

; LENGTH: 3292

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-320-878-22

Query Match 4.5%; Score 54.4; DB 3; Length 3292;  
Best Local Similarity 46.0%; Pred. No. 0.001;  
Matches 178; Conservative 1; Mismatches 208; Indels 0; Gaps 0;

QY 51 GGCTGCGGCTGCGCGCGCGCGCCAAATACGACGCTACAGCTTCGGAGCTTCCC 110  
Db |||||  
QY 2849 GGCGCGCTGCTGCGGATCCGGCTCGNCCACCTGGACAGCTGGAACGGCGGAGTCCGC 2908  
Db |||||  
QY 111 ACGGACGAGCTGATGCGGCTGAGTCGGCTTACCGGACCGGCTGGACAAGTACAGCGG 170  
Db |||||  
QY 2909 GCTGCGCGGAGTACCTCTCCGGGCTCGCGGATCGCCCGCATCGGCCTTCGCGGTGAC 2968  
Db |||||  
QY 171 CGAGCACTGGGCGGAGAGCTGAGTACCTGAGATCAGCTTCGCGCTGCAACCGCTTGTCT 230  
Db |||||  
QY 2969 CGCGCCGACACCGACCGGCTTGTGCACTTTCACCTGCGCACCGAGCGCGCGGCA 3028  
Db |||||  
QY 231 GCGGACACGAGGCGCTTGTGCACTTGTGCACTTGTGCACTTGTGCACTTGTGCACTTGTG 290  
Db |||||  
QY 3029 GCTGCGGAGCACCTTCGACGCGCGCGGATCGACACCTTCACGACTACCGCGGTACCGT 3088  
Db |||||  
QY 291 CGCGGCGCTCGCGAGCTATCCGAGCTGCGCTTTCGGGGGCTGCTGCGCGCGCGCA 350  
Db |||||  
QY 3089 GCACCTCTCGCGCGCTAGCGGGGAGGACCGCGCGGAAAGGCTGCTTCCCGGGCGCA 3148  
Db |||||  
QY 351 CTGCTCAAGCGCTGCAAGCAGGCGCTTCAGGCTTCGGCAGTCCCGAGCCGCGCGCA 410  
Db |||||  
QY 3149 GAGCTTCGCGGCGGAGGCTTCTCAGCTTCGCGATCGGCCCGCACCTGGAGGCGCGGAGGC 3208  
Db |||||  
QY 411 GGTGCTGCGGAGCTTCAGCGCGCGCA 437  
Db |||||

```
Db      3209 GCTGCGGGTGTATCGACGCCCTGCGCGA 3235
|||||
RESULT 15
US-09-141-908-15
; Sequence 15, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION: Gary
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; FILE OF INVENTION: Modular PKS Gene Cluster as Scaffold
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 15
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-15

Query Match      4.5%; Score 54.4; DB 4; Length 3292;
Best Local Similarity 46.0%; Pred. No. 0.0011;
Matches 178; Conservative 1; Mismatches 208; Indels 0; Gaps 0;

QY      51  GGCTGTGGGCTGCTGCGCGCGCGGCCCAATACGAACGCTACAGTTCCGCGAGCTTCCC 110
Db      2849  GGCGCGCGTGTCTCGGATCGGCTCGNCCACTTGGACACGCGCGCAGGTGGC 2908

QY      111  ACGGACAGAGCTGATCCGCTCGAGTCGGCTTACCGGCACGCGCTGGACAAGTACAGCGG 170
Db      2909  GCTGGCGCGGAGTACTCTCCGGGCTCGCGGACTGCGCGCATCGGCTCCCGGTGAC 2968

QY      171  CGAGCACTGGCGCGAGCSTKGGCTACCTGGAGATCAGCTGCGGCTGCACCGTTGCT 230
Db      2969  CGGCGCCGACACCGACCCCGGTCTGGCACCTCTTCACCGTGCGCACCGAGCGCGCGACGA 3028

QY      231  GCGCGACAGCGAGGCTTCTGCCACCGCAACTGCAAGCGCGCGCGCGAGCCGAGCCCGC 290
Db      3029  GCTGGCAGCCACCTCGACGCGCGCGGATCGACACCTCTACGCACTACCGGTACCGGT 3088

QY      291  CGCGGCGCTTGGCAGTATCCGAGTTCGGCTCTTTCGGGGGCTGTGCGCGCGCGCA 350
Db      3089  GCACCTCTCGCCCGCTACGCGGCGAGGCACCGCGGAAGGCTCGCTCCCGCGGCGCA 3148

QY      351  CTGCGCTCAAGCGGTGCAAGCAGGCGCTGCCAGCTTCGCGCAGTCCCGAGCCAGCGCGCA 410
Db      3149  GAGCTTCGCGCGCGAGGTCTCTAGCTGCGGATCGGCGCGGACCTGGAGCGCGCGCAGGC 3208

QY      411  GGTGTGGCGGACTTCCAGCGCGCGA 437
Db      3209  GCTGCGGGTGTATCGACGCCGTGCGCGA 3235
```

Search completed: January 30, 2004, 18:32:24  
Job time : 69.8238 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 07:55:54 ; Search time 2524.98 Seconds  
(without alignments)  
11579.600 Million cell updates/sec

Title: US-09-729-674-1\_COPY\_63\_1265

Perfect score: 1203

Sequence: 1 atggagcggggcgccgggg.....tggaaactggagagaccagc 1203

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*  
29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1033.2	85.9	1201	9	AL545483
2	977.2	81.2	1201	9	AL514488
3	969.8	80.6	1201	9	AL556976
4	957.8	79.6	1201	13	BX425905

5	954	79.3	1148	13	BU902192
6	945.8	78.6	1075	9	AL546910
7	929.6	77.3	1201	9	AL541167
8	923.8	76.8	1656	11	AK047506
9	923	76.7	1201	9	AL517455
10	912.6	75.9	1652	11	AK017797
11	866.8	72.1	1201	9	AL543090
12	863.6	71.8	1201	9	AL554750
13	838	69.3	939	13	BX386576
14	834	69.3	1201	9	AL558186
15	822.2	68.3	1152	9	AL550966
16	805.8	67.0	1040	9	AL558966
17	801	66.6	846	13	BX38140
18	787.6	65.5	944	13	BX56662
19	784.4	65.2	1109	12	BM550012
20	771.4	64.1	774	9	AL598447
21	766.6	63.7	1083	12	BM920687
22	765.6	63.6	1201	13	BX387779
23	761.2	63.3	846	12	B1253401
24	760.4	63.2	899	13	BQ437797
25	756.4	62.9	1200	13	BU902214
26	752.2	62.5	904	13	BU931550
27	746.2	62.0	962	14	CD516948
28	740.4	61.5	878	13	BUI83287
29	737	61.3	827	12	B1760850
30	735.4	61.1	1103	13	BQ233085
31	726	60.3	875	14	CD359485
32	720.6	59.9	1430	11	AK007624
33	718	59.7	1201	13	BX437872
34	717.2	59.6	912	13	BUI175306
35	711.8	59.2	945	13	BQ677531
36	704	58.5	782	10	B8877856
37	702.6	58.4	1088	10	BQ258933
38	702.4	58.4	837	13	BQ428356
39	700.8	58.3	1201	13	BX460044
40	700.2	58.2	707	10	BG747825
41	698	58.0	846	13	BUI189067
42	697.8	58.0	713	10	B2746070
43	691.6	57.5	887	13	BQ718961
44	680.2	56.5	943	13	BX539494
45	674.8	56.1	949	10	BG323659

## ALIGNMENTS

RESULT 1  
AL545483  
LOCUS  
DEFINITION  
AL545483 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0D1015K22 5-PRIME, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1201)  
Li W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
On Feb 15, 2001 this sequence version replaced gi:12877964.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
Bp 191 91006 EVRY cedex - France  
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 2233.f For more information about this cluster, see  
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D1015Bf11QPI&cluster=2233.f. Contact : Feng Liang Email : fliang@lifetech.com URL :



089

Query Match	81.2%	Score	977.2	DB	9	Length	1201	
Best Local Similarity	94.0%	Pred. No.	8.5e-222					
Matches 102;	Conservative	31	Mismatches	30	Indels	4	Gaps	4

Qy	1	ATGGAGCCGGGGCGCCGGGGGGGGCGCGCGCTGCTAGCGCTGCTGTGCGTGCCTTCGGCG	60
Db	62	ATGGAGCCGGGGCGCCGGGGGGGGCGCGCGCTGCTAGCGCTGCTGTGCGTGCCTTCGGCG	121
Qy	61	CTGCGCGCCGGCGCGCCCAATAGCAACGGCTACAGCTTCCGACAGTTCCCACGGACGAG	120
Db	122	CTGGCGCGCGGGCGGNCNAATACAAACGGCTACAGHTTCCGCAGCTTCCCAACGGACGAG	181
Qy	121	CTGATGCGCTCGAGTTCGGCTACCGACCGCTCGCAAGTACAGCGGCGACGACTGG	180
Db	182	CTGATGCGCTCGAGTTCGGCTACCGACCGCTCGCAAGTACAGCGGCGACGACTGG	241
Qy	181	GCCGAGAGCTGGCTACCTCGAGATCAGCTCGGCTGCACCGCTTGTTCGGCGACAGC	240
Db	242	GCCGAGAGCTGGCTACCTCGAGATHAGCTCGGCTGCACCGCTTGTTCGGCGACAGC	301
Qy	241	GAGGCTTCTGCACCGCAACTGCAGCGCGCGCGCAGCCGAGCGCGCGCGCGCGCTC	300
Db	302	GAGGCTTCTGTACCGCAACTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCTC	361
Qy	301	GCCAGTATCCGAGCTCGGCTCTTCGGGGGCTTGTGGCGCGCGCGCGCTTCAG	360
Db	362	GCCAGTATTCGAGCTCGCTCTTCGGGGGCTTGTGGCGCGCGCGCGCTTCAG	420
Qy	361	CGCTCAAGCAGGGCTCGCCAGCTTCGGCAGTCCCGACCGCAGCGCGCGCGCGCTGCGC	420
Db	421	CGCTCAAG-AGGGCTCTCMAGCHTYYGCHAGTWCWAGCHGCGAGGCTGCTGCGC	479
Qy	421	GACTTCAGCGCGCGAGCCCTCAAGTTCGTGAGTTCGCTTACTTCAGGCAAAAT	480
Db	480	GACTTTHAGCGCGCGAGCWCCTAAAAATTCYTGCAGTTCGCTTACTTAAAGAAAT	539
Qy	481	CTCCCCAAGCCATCGCGCTGCTCACACCTTTCTACTGAAGCATCTGATGACAAATG	540
Db	540	ATWCCCAARCWATCGCGCTGCTCAAACTTTTACTGAAGATCTCTGATGAAGAAATG	599
Qy	541	ATGAGAGAGACATGGCATATTTATAGAGCTTCGCTGCTCGGAGACTACATTAAAGAC	600
Db	600	ATGALAGAGACATGGNATATTAARACATCGTGGTGCTGAGGACTACATTAAAGAC	659
Qy	601	CTGGAAAACAAGTCATATGAAAGCTCTTTCATCGAGCAGTGGGGCATACAACGTTAG	660
Db	660	CTGGAAAACAAGTTATATGAAAGCWTGTTTCATCCGAGCAGTGGGGCATACAACGTTAG	719
Qy	661	AATCGAGAAATCCATCACAGACATGAGCTGGCTTCCCGACTTCTTCAAGACCTTT	720
Db	720	AACTCGAGAAACATCCATCACAGACATGAGCTGGCTTCCCGACTTCTTCAAGACCTTT	779
Qy	721	TACGAGTGTCTCGAGCTCGAGGGTTCAGGGAGATCAAGACTTCAAGATTTCTAC	780
Db	780	TACGAGTGTCTCGAGCTCGAGGGTTCAGGGAGATCAAGACTTCAAGATTTCTAC	839
Qy	781	CTTTCCATAGCAGATCAATATGTAGAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAC	840
Db	840	CTTTCCATAGCAGATCAATATGTAGAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAC	899
Qy	841	CTCACCCAGTTATAGGAGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTAC	900
Db	900	CTCACCCAGTTATAGGAGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTAC	959
Qy	901	TTGCGATTTTGCCTATTATTAAGTTGAAGCAGCTGAAGAATGCAGCCCCCTGTGCAGTCAGC	960
Db	960	TTGCGATTTTGCCTATTATTAAGTTGAAGCAGCTGAAGAATGCAG-CCCCGTGCGAGTCAGC	1018
Qy	961	TATCTGCTCTTTGATCAGAATGACAGGTATCGACGAACTCGTGGTATTATTCAGTAC	1020
Db	1019	TATCTGCTCTTTGATCAGAATGACAGGTATCGACGAACTCGTGGTATTATTCAGTAC	1078
Qy	1021	CACAGGACACTTGGGGCTCTCGGATGAGCACTTCCAGCCAGACCTCGAAGAGTTTCAG	1080
Db	1079	CACAGGACACTTGGGGCTCTCGGATGAGCACTTCCAGCCAGACCTCGAAGAGTTTCAG	1137

Qy 1081 TTCTTTA 1087  
 Db 1138 TTYTTAA 1144

RESULT 3  
 AL556976  
 LOCUS 1201 bp mRNA linear EST 31-MAY-2003  
 DEFINITION Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens cDNA  
 clone CSODH005YI01 5-PRIME, mRNA sequence.  
 ACCESSION AL556976  
 VERSION AL556976.2 GI:31278776  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12900137.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr  
 was not normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 2233.f. For more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CSODH005AE01QPI&cluster=2233.f. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CSODH005AE01QPI.  
 Location/Qualifiers  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODH005YI01"  
 /tissue\_type="T CELLS (JURKAT CELL LINE)"  
 /cell\_line="JURKAT CELL LINE"  
 /clone\_lib="Homo sapiens T CELLS (JURKAT CELL LINE)"  
 /notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."  
 272 a 339 c 320 g 234 t 36 others

BASE COUNT 272 a 339 c 320 g 234 t 36 others  
 ORIGIN

Query Match 80.6%; Score 969.8; DB 9; Length 1201;  
 Best Local Similarity 97.7%; Pred. No. 4.9e-220;  
 Matches 1015; Conservative 9; Mismatches 11; Indels 4; Gaps 4;

Qy 3 G G A C C G G G C G C G G G G G C C G C G C T G C T A G C G C T G C T G C G T G C G T G C G T G C G C G C T G C G C G C T 62  
 Db 58 G G A T C C G G G G C G G G G G G C C G C G C T G C T A G C G C T G C T G C G T G C G T G C G T G C G C G C T G C G C G C T 117

Qy 63 G C S C G C G G G C G C C C A A T A C G A A C G C T A C A G T T C C G A G C T T C C C A C G G A C G A G C T 122  
 Db 118 G C - C G C G G G C G G - C C A A T A C G A A C G C T A C A G T T C C G A G C T T C C C A C G G A C G A G C T 175

Qy 123 G A T G C G C T C G A G T C G G C C T A C C G G C A C G C G C T G G A C A A G T A C A G C G G C G A G C A C T G G G C 182  
 Db 176 G A T G C C G C T C G A G T C G G C C T A C C G G C A C G C G C T G G A C A A G T A C A G C G G C G A G C A C T G K G C 235

Qy 183 C G A G A G S T K G G C T A C T T G A G A T A C A G C T G G G C T G C A C C G C T T G C T G C G G A C A G C G A 242  
 Db 236 C G A G A G C T G G G C T A C T C G A G A T A C A G C T G G G C T G C A C C G C T T G C T G C G C A C A G C - A 294

Qy 243 G G C C T T C T G C A C C G G A A C T G A G C G C C G C G C G C A G C C G C G A G C C G C G C G C G C G C C T G C G 302

Db 295 GGCCTTCTGCCACCGCAAATGACAGGCGCGCGCGAGCCGCGCGCGCGCTCGC 354  
QY 303 CAGCTATCCCGAGCTGCGCTCTTTCGGGGCTGTGGCGCGCGCGCGCGCGCTCGC 362  
Db 355 CAGCTATCCCGAGCTGCGCTCTTTCGGGGCTGTGGCGCGCGCGCGCGCTCGC 414  
QY 363 CTGCAAGAGGCGCTGCGAGCTTTCGGCGAGTTCGCGCGCGCGCGCGCGCTCGC 422  
Db 415 CTGCAAGAGGCGCTGCGAGCTTTCGGCGAGTTCGCGCGCGCGCGCGCTCGC 474  
QY 423 CTTCCAGGCGCGCGAGCTTTCGAGTTCGCGAGTTCGCGAGTTCGCGAGTTCG 482  
Db 475 CTTCCAGGCGCGCGAGCTTTCGAGTTCGCGAGTTCGCGAGTTCGCGAGTTCG 534  
QY 483 CCCCAAGGCGCGCGCTGCTGCTCAACCTTCTTCTGAGGCGCGCGCGCGCTG 542  
Db 535 CCCCAAGGCGCGCGCTGCTGCTCAACCTTCTTCTGAGGCGCGCGCGCGCTG 594  
QY 543 GAAGAGGCGCGCGCTGCTGCTCAACCTTCTTCTGAGGCGCGCGCGCGCTG 602  
Db 595 GAAGAGGCGCGCGCTGCTGCTCAACCTTCTTCTGAGGCGCGCGCGCGCTG 654  
QY 603 GGAACCAAGTCAATGAAAGCTTTCATCCGAGCGTGGCGCGCGCGCGCTG 662  
Db 655 GGAACCAAGTCAATGAAAGCTTTCATCCGAGCGTGGCGCGCGCGCGCTG 714  
QY 663 CTGGAAGACATCCATCAGACATGAGCTGCGCTTCCGAGCTTCTTCAAGCTTT 722  
Db 715 CTGGAAGACATCCATCAGACATGAGCTGCGCTTCCGAGCTTCTTCAAGCTTT 774  
QY 723 CAGTGTCTCGAGCTCGAGGCTTCGAGGAGTTCAGGAGTTCAGGAGTTCACCT 782  
Db 775 CAGTGTCTCGAGCTCGAGGCTTCGAGGAGTTCAGGAGTTCAGGAGTTCACCT 834  
QY 783 TTCCATAGCAGATCAATGATGAGGCTTCGAGGAGTTCAGGAGTTCAGGAGTTC 842  
Db 835 TTCCATAGCAGATCAATGATGAGGCTTCGAGGAGTTCAGGAGTTCAGGAGTTC 894  
QY 843 CACCCAGCTATAGGAGCTTTCGAGGAGTTCGAGGAGTTCAGGAGTTCAGGAGTTC 902  
Db 895 CACCCAGCTATAGGAGCTTTCGAGGAGTTCGAGGAGTTCAGGAGTTCAGGAGTTC 954  
QY 903 CAGTGTCTCGAGCTCGAGGCTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCAG 962  
Db 955 CAGTGTCTCGAGCTCGAGGCTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCAG 1014  
QY 963 TCTGCTTTTCATCAGATGAGGCTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCAG 1022  
Db 1015 TCTGCTTTTCATCAGATGAGGCTTCGAGGAGTTCGAGGAGTTCGAGGAGTTCAG 1073  
QY 1023 CAGGACACTTGGGCGCTC 1041  
Db 1074 MAGGACACTTGGGCGCTC 1092

RESULT 4  
BX425905  
LOCUS BX425905 1201 bp mRNA linear EST 15-MAY-2003  
DEFINITION BX425905 Homo sapiens NEUROBLASTOMA Hom sapiens cDNA clone  
CSODA004YM07 5-PRIME, mRNA sequence.  
ACCESSION BX425905  
VERSION BX425905.1 GI:30774508  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS Li W.B., Gruber C., Jessee J., and Polayes D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
CONTACT Genoscope  
COMMENT Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by life technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 2233.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSOAA004AG04Q1&cluster=2233.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CSOAA004AG04Q1.

FEATURES  
source

Location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODA004YM07"  
/tissue\_type="NEUROBLASTOMA"  
/clone\_lib="Homo sapiens NEUROBLASTOMA"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."  
BASE COUNT 250 a 334 c 314 g 248 t 55 others  
ORIGIN

Query Match 79.6%; Score 957.8; DB 13; Length 1201;  
Best Local Similarity 93.7%; Pred. No. 3.5e-217;  
Matches 1019; Conservative 28; Mismatches 31; Indels 9; Gaps 5;  
QY 1 ATGGAGCGGGCGCGCGGGGGCGCGCGCTGTAGCGCTGTGTGGTGGCTTGGCG 60  
Db 37 ATGGAGCGGGCGCGCGGGGGCGCGCGCTGTGTGGTGGCTTGGCG 96  
QY 61 CTGGCGCGCGGGCGCGCGCAATACGAAAGCTACAGCTTCCGAGCTTCCACGGGACGAG 120  
Db 97 CTGGCGCGCGGGCGCGCGCAATACGAAAGCTACAGCTTCCGAGCTTCCACGGGACGAG 155  
QY 121 CTGATGCGCGCTGAGTGGCGCTTACCGGACGCGCTGGACAAGTACAGCGGAGCACTGG 180  
Db 156 CTGATGCGCGCTGAGTGGCGCTTACCGGACGCGCTGGACAAGTACAGCGGAGCACTGG 215  
QY 181 GCGGAGAGCTTGGCTTACTGTGAGATCAGCTTGGGCTGCGCGCTTCTGCGGACAGC 240  
Db 216 GCGGAGAGCTTGGGCTTACTGTGAGATCAGCTTGGGCTTCTGCGGACAGC -- 273  
QY 241 GAGGCTTCTGCGCAACGCAACTGCGCGCGCGCGCGAGCGCGCGCGCGCGCTC 300  
Db 274 SAGGCTTCTGCGCAACGCAACTGCGCGCGCGCGCGAGCGCGCGCGCGCGCTC 333  
QY 301 GCGAGCTATCCGAGCTGGGCTCTTTCGGGGGCTGTGTGGCGCGCGCTGCTGCTCAAG 360  
Db 334 GCGAGCTATCCGAGCTGGGCTCTTTCGGGGGCTGTGTGGCGCGCGCTGCTGCTCAAG 393  
QY 361 CGCTCAAGAGCGGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAG 420  
Db 394 CGCTCAAGAGCGGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAG 453  
QY 421 GACTTCCAGCGCGCGAGCGCTTCAAGTCTCTGAGTTCGCTTACTTCAAGGCAATAAT 480  
Db 454 GACTTCCAGCGCGCGAGCGCTTCAAGTCTCTGAGTTCGCTTACTTCAAGGCAATAAT 513  
QY 481 CTCCCAAGGCGCATCGCGCTGCTTCAACCTTCTTCTGAGGATCTCTGATGAGCAATG 540  
Db 514 CTCCCAAGGCGCATCGCGCTGCTTCAACCTTCTTCTGAGGATCTCTGATGAGCAATG 573  
QY 541 ATGAAGAGGCAATGGCATATTAAGAGCGCTGCTGGTGGCGGAGGACTCATTAAGAG 600  
Db 574 ATGAAGAGGCAATGGCATATTAAGAGCGCTGCTGGTGGCGGAGGACTCATTAAGAG 633  
QY 601 CTGAAACCAAGTCAATATGAAGCGCTTTCATCCGAGCAGTGGGGGCGATACAGGGTGA 660  
Db 634 CTGAAACCAAGTCAATATGAAGCGCTTTCATCCGAGCAGTGGGGGCGATACAGGGTGA 693

```
QY 661 AACTGGAGAAATCCATCACAGACATGAGGTGGCCCTTCCGACTTCTTCAAAAGCCTTT 720
Db 694 AACTGGAGAAATCCATCACAGACATGAGGTGGCCCTTCCGACTTCTTCAAAAGCCTTT 753
QY 721 TACGAGTGTCTCGAGAGCTCGAGGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTAC 780
Db 754 TACGAGTGTCTCGAGAGCTCGAGGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTAC 813
QY 781 CTTTCCATAGAGATCAATATGTAGAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAAC 840
Db 814 CTTTCCATAGAGATCAATATGTAGAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAAC 873
QY 841 CTCACCCAGATTAGAGAGCTATCCGGTTGAGAAATTTGGTACCATGATATCATAC 900
Db 874 CTCACCCAGATTAGAGAGCTATCCGGTTGAGAAATTTGGTACCATGATATCATAC 933
QY 901 TTGCAGTTTGCTATTATTAAGTTGAAGACCTGGAAGATGAGCCCTCTGAGTGCAGC 960
Db 934 TTGCAGTTTGCTATTATTAAGTTG-ACGACCTGAAGATGAGCCCTCTGKAGTGCAGC 992
QY 961 TATCTGCTTTGATGAGATGACAGGTGATGAGAGAACCTGGTGTATACCATGATAC 1020
Db 993 TATCTGCTTTGATGAGATGACAGGTGATGAGAGAACCTGGTGTATACCATGATAC 1051
QY 1021 CACAGGAGACATCTGGGCTCTCGGATGAGCACTTCCAGCCCGAGACCTGAGGAGTTCAG 1080
Db 1052 MARGG---AMATGGGGSYTYGGTGANVYTYGACCCARACTTAACARTYAMMTYTT 1107
QY 1081 TTCTTTTA 1087
Db 1108 TTWTTTW 1114

RESULT 5
LOCUS BU902192
DEFINITION BU902192 1148 bp mRNA linear EST 17-OCT-2002
5', mRNA sequence.
AGENCOURT 10127451 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6502891
BU902192
BU902192.1 GI:24084105
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
UNPUBLISHED
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M14059 row: k column: 20
High quality sequence stop: 754.
Location/Qualifiers
1..1148
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6502891"
/tissue_type="leiomysarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-Sport6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

FEATURES
source
```

```
BASE COUNT 253 a 363 c 315 g 216 t 1 others
ORIGIN
Query Match 79.3%; Score 954; DB 13; Length 1148;
Best Local Similarity 96.2%; Pred. No. 2.8e-216;
Matches 1008; Conservative 2; Mismatches 32; Indels 6; Gaps 3;
QY 5 AGCCGGGGCGCCGGGGGGCGGCGCTGCTAGCGCTGCTGTGGCTCGGCTCGGCGCTGC 64
Db 1 AGCCGGGGCGCCGGGGGGCGGCGCTGCTAGCGCTGCTGTGGCTCGGCGCTGC 60
QY 65 GGGCCGGGGCGCCCGCCAAATACGAACGCTACAGTTCCGAGCTTCCACGGGACGAGCTGA 124
Db 61 GGGCCGGGGCGCCCGCCAAATACGAACGCTACAGTTCCGAGCTTCCACGGGACGAGCTGA 120
QY 125 TGCCTGCTCAGTTCGCTTACCGGCACGCGCTGCGCAAGTACAGCGCGGAGCACTGGGCGG 184
Db 121 TGCCTGCTCAGTTCGCTTACCGGCACGCGCTGCGCAAGTACAGCGCGGAGCACTGGGCGG 180
QY 185 AGAGCTKGGTACCTGAGATCAGCTTCGCGCTTCACCGCTTGTGCGCGACACGAGG 244
Db 181 AGAGCTKGGTACCTGAGATCAGCTTCGCGCTTCACCGCTTGTGCGCGACACGAGG 240
QY 245 CTTCTGCGACCGCAACTGCGCGCGCGCGCGAGCCGCGCGCGCGCGCGCGCGCGCGCG 304
Db 241 CTTCTGCGACCGCAACTGCGCGCGCGCGCGAGCCGCGCGCGCGCGCGCGCGCGCGCG 300
QY 305 GCTATCCGAGTGGCGCTCTTTCGGGGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCT 364
Db 301 GCTATCCGAGTGGCGCTCTTTCGGGGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCT 360
QY 365 GCAAGCAGGGCTGCGAGCTTCGCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGAGCT 424
Db 361 GCAAGCAGGGCTGCGAGCTTCGCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGAGCT 420
QY 425 TCCAGCGCGCGAGCGCTTACAGTTCTCGAGTTCGCTTACTTCAAGGCAAAATATCTCC 484
Db 421 TCCAGCGCGCGAGCGCTTACAGTTCTCGAGTTCGCTTACTTCAAGGCAAAATATCTCC 480
QY 485 CCAAGCCATGCGCGCTGCTCAGCTTCTTCTGAGAGCATCTCTGATGAGAAATGATGA 544
Db 481 CCAAGCCATGCGCGCTGCTCAGCTTCTTCTGAGAGCATCTCTGATGAGAAATGATGA 540
QY 545 AGAGAAATGCGCATATTATTAAGAGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 604
Db 541 AGAGAAATGCGCATATTATTAAGAGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 600
QY 605 AAACCAAGTCATATGAAAGCCTGTTTCATCCGAGCAGTGGCGGCATCAACGGTGAGAACT 664
Db 601 AAACCAAGTCATATGAAAGCCTGTTTCATCCGAGCAGTGGCGGCATCAACGGTGAGAACT 660
QY 665 GGAGAACATCCATCACAGACATGAGCTGGCCCTTCCGACTTCTTCAAAGCCTTTTACG 724
Db 661 GGAGAACATCCATCACAGACATGAGCTGGCCCTTCCGACTTCTTCAAAGCCTTTTACG 720
QY 725 AGTGTCTCGAGCTGCGAGGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTACCTTT 784
Db 721 AGTGTCTCGAGCTGCGAGGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTACCTTT 780
QY 785 CCAATAGCATCATATGTAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAACTCA 844
Db 781 CCAATAGCATCATATGTAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAACTCA 840
QY 845 CCCAGTTATAGGAGCTATCCGCTTGAGAAATTTGGCTACCATGATCATCTACTTGC 904
Db 841 CCCAGTTATAGGAGCTATCCGCTTGAGAAATTTGGCTACCATGATCATCTACTTGC 900
QY 905 AGTTTGCTATTATAGTTGAACGACCTGAAGATGCGAG-CCCCCTGTGCGAGTCAGTAT 963
Db 901 AGTTTGCTATTATAGTTGAACGACCTGAAGATGCGAG-CCCCCTGTGCGAGTCAGTAT 960
QY 964 CTGCTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1021
```

```

Db 961 CGGCTCTTTGATCAGAAAGACAGGCTCTGACGCAACAACTGGGGGTATTACCGCTCC 1020
QY 1022 ACAGGGACAC---TTGGGGCTCTTCGGA 1046
Db 1021 CCCCGGGACCCCTTTGGGGCCCTTCGGA 1048

RESULT 6
AL546910 1075 bp mRNA linear EST 31-MAY-2003
AL546910 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0DI032YH24 5-PRIME, mRNA sequence.
ACCESSION AL546910
VERSION AL546910.2 GI:31268743
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1075)
JOURNAL Li W.B., Gruber,C., Jessee,J. and Polayes,D.
COMMENT Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12880487.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2233.f For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DI032YH24&cluster=2233.f. Contact :
Feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DI032DD12QF1.
Location/Qualifiers
1..1075
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI032YH24"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="First strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 240 a 317 c 256 g 212 t 10 others
ORIGIN

Query Match 78.6%; Score 945.8; DB 9; Length 1075;
Best Local Similarity 98.8%; Pred. No. 2.4e-214;
Matches 966; Conservative 6; Mismatches 4; Indels 2; Gaps 2;

QY 2 TGGAGCGGGGCGCGGGGGCGCGGGCGCTGTAGCGCTGTGTGGTGGCTTGGCGCGC 61
Db 62 TGGAGCGGGGCGCGGGGGCGCGGGCGCTGTAGCGCTGTGTGGTGGCTTGGCGCGC 121
QY 62 TGGCGCGGGGCGCGGGCGCGCGCGCTGTAGCGCTGTGTGGTGGCTTGGCGCGC 121
Db 122 TGGCGCGGGGCGCGGGCGCGCGCGCTGTAGCGCTGTGTGGTGGCTTGGCGCGC 181
QY 122 TGATCGCGCTGTAGCTGGCGCTACCGGACCGCTGGACAAAGTACAGCGGAGCACTGGG 181
Db 182 TGATCGCGCTGTAGCTGGCGCTACCGGACCGCTGGACAAAGTACAGCGGAGCACTGGG 241
QY 182 CCGAGAGCTKGGCTACTGGAGATCAGCTGGCTGGCTGGACCGCTGCTCGCGGACAGCG 241
Db 242 CCGAGAGCTGGGTACTGTGAGATCAGCTGGCTGGCTGGACCGCTGCTCGCGGACAGCG 301
QY 242 AGGCTTTCTGCCACCGCAACTGTCAGCGCGCGCGCGCGAGCCCGCGCGCGCGCTCG 301

```

```

Db 302 AGGCTTTCTGCCACCGCAACTGTCAGCGCGCGCGCGCGAGCCCGCGCGCGCTCG 361
QY 302 CAGCTATCCCGAGCTTGGCGCTTTGGGGGGCTGTGTGGCGCGCGCTGTGTGGCGCACTGCTCAAGC 361
Db 362 CAGCTATCCCGAGCTTGGCGCTTTGGGGGGCTGTGTGGCGCGCGCTGTGTGGCGCACTGCTCAAGC 421
QY 362 GCTGCAAGCAGGCGCTTGGCGCGCTTCCGCGAGTCCCGAGCCCGCGCGCGAGGTGTGTGGCGG 421
Db 422 GCTGCAAGCAGGCGCTTGGCGCGCTTCCGCGAGTCCCGAGCCCGCGCGCGAGGTGTGTGGCGG 481
QY 422 ACTTCAGCGCGCGCGAGCCCTACAAGTTCCTCGAGTTCGCTTACTTCAAGGCAAAATAATC 481
Db 482 ACTTCAGCGCGCGCGAGCCCTACAAGTTCCTCGAGTTCGCTTACTTCAAGGCAAAATAATC 541
QY 482 TCCCAAGCCATCGCGCGCTGCTCACACCTTTCTACTGAAGCATCTGTATGACGAAATGA 541
Db 542 TCCCAAGCCATCGCGCGCTGCTCACACCTTTCTACTGAAGCATCTGTATGACGAAATGA 601
QY 542 TGAAGAGCAACATGGCATATTATAAGAGCCTCCCTGGTGCAGGACTACATTAAAGACC 601
Db 602 TGAAGAGCAACATGGCATATTATAAGAGCCTCCCTGGTGCAGGACTACATTAAAGACC 661
QY 602 TGAAGAGCAACATGGCATATTATAAGAGCCTCCCTGGTGCAGGACTACATTAAAGACC 661
Db 662 TGAAGAGCAACATGGCATATTATAAGAGCCTCCCTGGTGCAGGACTACATTAAAGACC 721
QY 662 ACTGGAGAACATCCATCAGACATGAGAGCTGGCGCTTCCCGACTTCTTCAAGGCTTTT 721
Db 722 ACTGGAGAACATCCATCAGACATGAGAGCTGGCGCTTCCCGACTTCTTCAAGGCTTTT 781
QY 722 AGAGTGTCTCGACGCTTCCGAGGCTTCCAGGAGATCAAGGACTTCAAGGATTTCTACC 781
Db 782 AGAGTGTCTCGACGCTTCCGAGGCTTCCAGGAGATCAAGGACTTCAAGGATTTCTACC 841
QY 782 TTTCCATAGCAGATCATTTATGTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAGAAC 841
Db 842 TTTCCATAGCAGATCATTTATGTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAGAAC 901
QY 842 TCACCCCGAGTTATAGAGGCTATCCGGTTGAGAAATTTGTGGTACCATGTATCATTTACT 901
Db 902 TCACCCCGAGTTATAGAGGCTATCCGGTTGAGAAATTTGTGGTACCATGTATCATTTACT 961
QY 902 TGCAGTTTGCCTATTATTAAGTTGAACGACCTTGAAGATGCAAGTGCAGCCCGCTGTGAGTCACT 961
Db 962 TGCAGTTTGCCTATTATTAAGTTGAACGACCTTGAAGTGCAGTGCAGTGCAGTGCAGT 1019
QY 962 ATCTGCTCTTTGATCAGA 979
Db 1020 ATCTGCTCTTTGATMRAA 1037

```

```

RESULT 7
AL541167 1201 bp mRNA linear EST 12-MAY-2003
LOCUS AL541167 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE005YF12
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION AL541167
VERSION AL541167.2 GI:30545079
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1201)
JOURNAL Li W.B., Gruber,C., Jessee,J. and Polayes,D.
COMMENT Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12871971.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of

```

Invitrogen. This sequence belongs to sequence cluster 2233.f. For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DE005DC06P1&cluster=2233.f>. Contact : Feng Liang Email : [liang@lifetech.com](mailto:liang@lifetech.com) URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DE005DC06P1

## FEATURES

```

location/Qualifiers
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB005YF12"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT6; 1st strand cDNA
with a NotI-oligo(dT) primer. Five prime
double-strand cDNA was digested with NotI
the Not I and EcoRV sites of the pCMVSPORT6
Library was not normalized."
258 a 335 C 313 G 237 T 58 others
BASE COUNT
ORIGIN

```

Query Match	77.3%	Score 929.6	DB 9	Length 1201
Best Local Similarity	98.2%	Pred. No. 1.8e-210		
Matches 970	Conservative 10	Mismatches 4	Indels 4	Gaps 4
Qy	1	ATGAGAGCCGGGCGCCGGGGGCGCGCGCTGCTAGCGCTGTGTGCGTGGCTGGCTGGCGG	60	
Db	64	ATGAGAGCCGGGCGCCGGGGGCGCGCGCTGCTAGCGCTGTGTGCGTGGCTGGCTGGCGG	123	
Qy	61	CTGCGCGCCGGGCGCGGCCCAATACGAACGCTACAGCTTCCGCAGCTTCCACCGGACGAG	120	
Db	124	CTGCGCGCGGGCGC - CCCAATAACGAACGCTACAGCTTCCGCAGCTTCCACCGGACGAG	182	
Qy	121	CTGATGCGGCTCGAGTCGGCTACCGCAGCGCGCTGGACAAAGTACAGCGCGCAGACTG	180	
Db	183	CTGATGCGGCTCGAGTCGGCTACCGCAGCGCGCTGGACAAAGTACAGCGCGCAGCCTG	242	
Qy	181	GCCGAGAGCCTKGCTACCTCGGAGTACAGCCTGCGCTGCAACGCTGTGTGCGCGACAGC	240	
Db	243	GCCGAGAGCCTGGGCTACCTGGAGTACAGCCTGCGGCTGCAACGCTGTGTGCGCGACAGC	302	
Qy	241	GAGGCTTCTGCCACACCGCAACTGAGCGCGCGCGCAGCCGAGCCGCGCGGCTC	300	
Db	303	-AGGCTTCTGCCACACCGCAACTGAGCGCGCGCGCAGCCGAGCCGCGCGGCTC	361	
Qy	301	GCCAGCTATCCCGAGTGGCCCTCTTCGGGGGCGCTGCTGGCGCGCGGCACTGCCTCAAG	360	
Db	362	GCCAGCTATCCCGAGTGGCCCTCTTCGGGGGCGCTGCTGGCGCGCGGCACTGCCTCAAG	421	
Qy	361	CGTGTCAAGCAGGGCTGCGAGCCTTCCGCGAGTCCCGCAGCCAGCGCGAGTGTGTGGG	420	
Db	422	CGTGTCAAGCAGGGCTGCGAGCCTTCCGCGAGTCCCGCAGCCAGCGCGAGTGTGTGGG	481	
Qy	421	GACTTCCAGCGCGCGAGCCCTACAGTTCCTGCACTTCGCTTACTTCAAGGCAATAAT	480	
Db	482	GACTTCCAGCGCGCGAGCCCTACAGTTCCTGCACTTCGCTTACTTCAAGGCAATAAT	541	
Qy	481	CTCCCCAAGCCATCGCCGCTGCTCACACCTTCTACTGAAGCATCCTGATCAGCAATG	540	
Db	542	CTCCCCAAGCCATCGCCGCTGCTCACACCTTCTACTGAAGCATCCTGATGAATG	601	
Qy	541	ATGAAGAGGAACATGCGATATTATTAAGAGCGCTGCTGGTGGCGAGACTACATTAAAGAC	600	
Db	602	ATGAAGAGGAACATGCGATATTATTAAGAGCGCTGCTGGTGGCGAGACTACATTAAAGAC	661	
Qy	601	CTGGAAACCAAGTCATATGAAGCCTGTTCATCCGAGCAGTGCGGGCATACACGGTGAG	660	
Db	662	CTGGAAACCAAGTCATATGAAGCCTGTTCATCCGAGCAGTGCGGGCATACACGGTGAG	721	
Qy	661	AACTGTGAGAACATTCATACAGACATGGAGCTGGCCCTTCCGCACTTCTTCAAGCCCTTT	720	

Db	722	AATGAGAGACATCCATATCAGACATGGAGCTGGCCCTTCCGACTCTTTCAAAGCCTTT	781
Qy	721	TACGAGTGTCTCGCAGCCTCGCAGGGTTC CAGGAGATCAAGGACTTCAAGGATTTCTAC	780
Db	782	TACGAGTGTCTCGCAGCCTCGCAGGGTTC CAGGAGATCAAGGACTTCAAGGATTTCTAC	841
Qy	781	CTTTCCATAGCAGATCATTTATGTAGAAGTTCTGGAATGGAATAACAGTGTGAAGAGAAC	840
Db	842	CTTTCCATAGCAGATCATTTATGTAGAAGTTCTGGAATGCACATACAGTGTGAAGAGAAC	901
Qy	841	CTCACCCCAAGTTATAGSAGGCTATCCGGTTTGAGAAATTTGTGGCTACCATGTATCAATTAC	900
Db	902	CTCACCCCAAGTTATAGSAGGCTATCCGGTTTGAGAAATTTGTGGCTACCATGTATCAATTAC	961
Qy	901	TTGCACTTTTGCCTATTATTAAGTTGACGACCTTGACAAATGCAGCCCCCTGTGCAGTCAGC	960
Db	962	TTGCACTTTTGCCTATTATTAARITG-AMGACCTKGARAATGCAG-CCCTCTGTGCAGTCAGC	1019
Qy	961	TATCTGCTCTTTTGATCAGAATGACAAGG	988
Db	1020	TATYTSCTCTTTGWTAGAAATGADAGG	1047

RESULT 8	
AK047506	
LOCUS	
DEFINITION	

LOCUS	AK047506	1656 bp	mRNA	linear	HTC 05-DEC-2002
DEFINITION	Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:B930079M13 product:cartilage associated protein, full insert sequence.				

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANIS

VERSION	AK047506.1	GI:26338909
KEYWORDS	HTC; CAP trapper.	
SOURCE	Mus musculus	(house mouse)
ORGANISM	Mus musculus	

ORGANISATION

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS

**AUTHORS** Carninci, P. and Hayashizaki, Y.  
**TITLE** High-efficiency full-length cDNA cloning  
**JOURNAL** Meth. Enzymol. 303, 19-44 (1999)  
**MEDLINE** 99279253  
**PUBMED** 10349636

## REFERENCES

AUTHORS	Carninci, P., Shibata, Y., Hayatsu N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159

PUBMED  
REFERENCE

**AUTHORS** Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M., Suni, N., Ishi, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kiru, A. and Hayashizaki, Y.

**TITLE** RIKEN integrated sequence analysis (RISA) system—384-format sequencing pipeline with 384 multicapillary sequencer

**JOURNAL** Genome Res. 10(11), 1757-1771 (2000)

JOURNAL  
MEDLINE  
RETURNED

PUBMED REFERENCE	AUTHORS
11076861	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Arzawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakata, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Aoshimizu, M., Batalov, S., Casavant, T., Frieschmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.



Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gusticich, S., Hill, D.,  
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,  
Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,  
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.  
and Hayashizaki, Y.

**TITLE**  
Functional annotation of a full-length mouse cDNA collection

**JOURNAL**  
Nature 409 (6821), 685-690 (2001)

**MEDLINE**  
21085660

**PUBMED**  
11217851

**REFERENCE**

**AUTHORS**

**TITLE**  
The FANTOM Consortium and the RIKEN Genome Exploration Research

**JOURNAL**  
Group Phase I & II Team.

**AUTHORS**  
Analysis of the mouse transcriptome based on functional annotation

**REFERENCE**  
of 60,770 full-length cDNAs

**AUTHORS**  
Nature 420, 563-573 (2002)

**REFERENCE**  
6 (bases 1 to 1656)

**AUTHORS**  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,

**REFERENCE**  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,

**AUTHORS**  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,

**REFERENCE**  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kasugawa, T.,

**AUTHORS**  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

**REFERENCE**  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,

**AUTHORS**  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,

**REFERENCE**  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,

**AUTHORS**  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

**REFERENCE**  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,

**AUTHORS**  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,

**REFERENCE**  
Muramatsu, M., and Hayashizaki, Y.

**AUTHORS**  
Direct Submission

**REFERENCE**  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of

**AUTHORS**  
Physical and Chemical Research (RIKEN), Laboratory for Genome

**REFERENCE**  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

**AUTHORS**  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

**REFERENCE**  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,

**AUTHORS**  
URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222,

**REFERENCE**  
Fax: 81-45-503-9216)

**AUTHORS**  
cDNA library was prepared and sequenced in Mouse Genome

**REFERENCE**  
Encyclopedia Project of Genome Exploration Research Group in Riken

**AUTHORS**  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.

**REFERENCE**  
Division of Experimental Animal Research in Riken contributed to

**AUTHORS**  
prepare mouse tissues.

**REFERENCE**  
Please visit our web site for further details.

**AUTHORS**  
URL: <http://genome.gsc.riken.go.jp/>

**REFERENCE**  
URL: <http://fantom.gsc.riken.go.jp/>

**AUTHORS**  
Location/Qualifiers

**REFERENCE**  
1..1656

**AUTHORS**  
/organism="Mus musculus"

**REFERENCE**  
/mol\_type="mRNA"

**AUTHORS**  
/strain="C57BL/6J"

**REFERENCE**  
/db\_xref="FANTOM\_DB:B930079M13"

**AUTHORS**  
/db\_xref="taxon:10090"

**REFERENCE**  
/clone="B930079M13"

**AUTHORS**  
/tissue\_type="cerebellum"

**REFERENCE**  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"

**AUTHORS**  
/dev\_stages="10 days neonate"

**REFERENCE**  
70..1272

**AUTHORS**  
/note="unnamed protein product; cartilage associated

**REFERENCE**  
protein (MGI:1891221, GI:191922, evidence: BLASTN,

**AUTHORS**  
99%, match=1655)

**REFERENCE**  
putative"

**AUTHORS**  
/codon\_start=1

**REFERENCE**  
/protein\_id="BAC33076.1"

**AUTHORS**  
/db\_xref="GI:26338810"

**REFERENCE**  
/translation="WGPRSPPTALLVLCVCAPTPOGQYERYSPFRDELMPLE

**AUTHORS**  
SAYRHLDQYSGEHWASVGLVSLRLHRLRDSFAFCHRNCSAATPAPAPGASH

**REFERENCE**  
AELRFGVLRRKCKKQGLFQFQSPQSRVLAQFOOREFYFLQPAYFKANDL

**AUTHORS**  
PKAIAAHTYLLKPKDDMMKRNMEYKSLFGAEDHIKLETKSYSLFVRAVYNG

**REFERENCE**  
ENWRTSISDMELALPDFLKAIFYECIAACEGSRKIDPKDFVLSIADHYVEVLECKRC

EEETLPVIGGYPVEKFAVWYHYLFAYKLNDAKNAAPCAVSYLLPFDSDRVMQONL  
VYVOYHRDKWGLSDEHFQRPPEAFQFNVTTLQKELYDFAQEHLMDDDEGEVVEYVDD  
LLETSSA"

BASE COUNT 341 a 498 c 448 g 369 t

ORIGIN

Query Match 76.8%; Score 923.8; DB 11; Length 1656;

Best Local Similarity 86.0%; Pred. No. 5.2e-209;

Matches 1021; Conservative 2; Mismatches 164; Indels 0; Gaps 0;

Qy 13 CGCCGGGGGGCGGGCGCTGTAGCGCTGTCTGTGCGCTGCGCTGCGCTGCGCGGGG 72

Db 79 CGCAGCCCTACGGCCCGCTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 138

Qy 73 CGCGCCCAATACCAAGCTACAGCTTCCGAGCTTCCACGGGACGAGCTGATCGCGCTC 132

Db 139 CGCGGGCAGTACGAGCGCTACAGCTTCCGAGCTTCCGCGGGAGCAGCTGATCGCGCTC 198

Qy 133 GAGTCGGCTTACCGGACGCGCTTGGACAAGTACAGCGGCGAGCCTGGGCGCGAGAGCSTK 192

Db 199 GAGTCGGCTTACCGGACGCGCTTGGACAAGTACAGCGGCGAGCCTGGGCGCGAGAGCSTK 258

Qy 193 GGCTACCTGGAGATCAGCTTGGCGCTGCACTTCTGCTGCGGACGAGCGCTTCTGCG 252

Db 259 GGCTACCTGGAGATCAGCTTGGCGCTGCACTTCTGCTGCGGACGAGCGCTTCTGCG 318

Qy 253 CACGCAACTGACG 312

Db 319 CACGCAACTGACG 378

Qy 313 GAGTCGGCTTCTGGGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372

Db 379 GAGTCGGCTTCTGGGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 438

Qy 373 GGCTGCCAGCTTCCG 432

Db 439 GGCTGCCAGCTTCCG 498

Qy 433 CGCGAGCGCTACAGTCTCTGCGAGTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTG 492

Db 499 CGCGAGCGCTACAGTCTCTGCGAGTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTG 558

Qy 493 ATCGCGCTGCTCACACCTTTTCTGTAAGCATCTCTGTAAGCATCTCTGTAAGCATCTCT 552

Db 559 ATCGCTGGCTCACACCTTTTCTGTAAGCATCTCTGTAAGCATCTCTGTAAGCATCTCT 618

Qy 553 ATGCGATATATAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 612

Db 619 ATGCGATATATAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 678

Qy 613 TCATATGAAAGCGCTTTCATCCGAGCGAGTGGCGGCGATACACGGTGAGAACTGGAGAAC 672

Db 679 TCGTACGAGCGCTTTCATCCGAGCGAGTGGCGGCGATACACGGTGAGAACTGGAGAAC 738

Qy 673 TCCATCACAGACATGGAGCGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAG 732

Db 739 TCCATTCGACATGGAGCGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAGCTTCCGAG 798

Qy 733 GCAGCGCTGCGGCGCTTCCAGGAGATCAAGGACTTCAAGGACTTCAAGGACTTCAAGGACT 792

Db 799 GCTGCTGCGAGGGGTGCGGAGATCAAGGACTTCAAGGACTTCAAGGACTTCAAGGACT 858

Qy 793 GATCATATATAGAGTCTTGGAAATGCAAAATACAGTGTGAAGAGAACCTCACCCCGT 852

Db 859 GATCATATATAGAGTCTTGGAAATGCAAAATACAGTGTGAAGAGAACCTCACCCCGT 918

Qy 853 ATAGAGGGCTATCCGCTTGGAGAAATTTCTGGCTACCATGATCATCTTACGATTTGCC 912

Db 919 ATAGAGGGCTATCCGCTTGGAGAAATTTCTGGCTACCATGATCATCTTACGATTTGCC 978

Qy 913 TATTATAGTTGAAACGACTGAAGAAATGCAAAATACAGTGTGAAGAGAACCTCACCCCGT 972

Db 979 TATTATAGTTGAAACGACTGAAGAAATGCAAAATACAGTGTGAAGAGAACCTCACCCCGT 1038



QY 973 GATCAGATGACAGGCTATGACAGAGAACTGTGTATATACAGTACCAAGGACACT 1032  
 Db 1039 GACACAGATGACAGGCTATGACAGAGAACTGTGTATATACAGTACCAAGGACACT 1098  
 QY 1033 TGGGGCTCTCGATGAGCACTTCCAGCCAGAGCTGAAGAGTTCAGTTCTTTAATGTG 1092  
 Db 1099 TGGGGCTCTCGATGAGCACTTCCAGCCAGAGCTGAAGAGTTCAGTTCTTTAATGTG 1158  
 QY 1093 ACCACACTCCAGAGAGCTGTATGACATTTCTTAAGGAAATATATATGATGATGAG 1152  
 Db 1159 ACAGCGCTCCAGAGAGAGCTGTATGACATTTCTTAAGGAAATATATATGATGAG 1218  
 QY 1153 GGAGAGTGTGGAATATGATGATGACCTCTTGAACCTGGAGGAGAC 1199  
 Db 1219 GGAGAGTGTGGAATATGATGATGACCTCTTGAACCTGGAGGAGAC 1265

RESULT 9.  
 ALS17455  
 LOCUS  
 DEFINITION CS0DA003YF11 5-PRIME, mRNA sequence.  
 ACCESSION ALS17455  
 VERSION ALS17455.2 GI:30534775  
 KEYWORDS EST.  
 ORGANISM Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li W.B., Gruber, C., Jesse, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT On Feb 13, 2001 this sequence version replaced gi:12780948.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 2233.f For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DA003YF11&cluster=2233.f. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue, Genoscope sequence ID : CS0DA003YF11.  
 Location/Qualifiers  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DA003YF11"  
 /tissue\_type="NEUROBLASTOMA"  
 /clone\_lib="Homo sapiens NEUROBLASTOMA"  
 /note="Vector: pCMVSPORT\_6; 1st strand cDNA was primed  
 with a NotI-oligo(dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."  
 BASE COUNT 270 a 327 c 306 g 240 t 58 others  
 ORIGIN

Query Match 76.7%; Score 923; DB 9; Length 1201;  
 Best Local Similarity 97.3%; Pred. No. 6.8e-209;  
 Matches 956; Conservative 13; Mismatches 8; Indels 6; Gaps 3;  
 QY 6 GCGGGGCGCGGGGGCGGGCGGCTGTAGCGTGTGTGCGTGGCGCTGGCGCTGGC 65  
 Db 65 GCGGGGCGCGGGGGCGGGCGGCTGTAGCGTGTGTGCGTGGCGCTGGCGCTGGC 124  
 QY 66 GCGGGGCGCGGGGGCGGGCGGCTGTAGCGTGTGTGCGTGGCGCTGGCGCTGGC 125

Db 125 CCGGGCGG----CCATACGAAACGCTACAGCTTCCGCGAGCTTCCACGGGACAGCTGAT 181  
 QY 126 GCGGCTCGAGTGGCGCTTACCGGACCGCGCTGGACAAGTACAGCGGAGACACTTGGGCGGA 185  
 Db 182 GCGGCTCGAGTGGCGCTTACCGGACCGCGCTGGACAAGTACAGCGGAGACACTTGGGCGGA 241  
 QY 186 GAGGCTGGCTTACCTGGAGATCAGCTTGGCGCTGACCGCTTCTGCGGACAGCGGAGGC 245  
 Db 242 GAGGCTGGCTTACCTGGAGATCAGCTTGGCGCTGACCGCTTCTGCGGACAGCGGAGGC 299  
 QY 246 CTTCTGCCACCGCAACTGACG 305  
 Db 300 CTTCTGCCACCGCAACTGACG 359  
 QY 306 CTATCCGAGCTGGCGCTCTTTCGGGGGCGCTTTCGGGGGCGCTTTCGGGGGCGCTTTCGG 365  
 Db 360 CWTCCCGAGCTGGCGCTCTTTCGGGGGCGCTTTCGGGGGCGCTTTCGGGGGCGCTTTCGG 419  
 QY 366 CAAGCAGGCTTGGCGAGCTTTCGCCAGTCCCGAGTCCCGAGTCCCGAGTCCCGAGTCCCGAG 425  
 Db 420 CAAGCAGGCTTGGCGAGCTTTCGCCAGTCCCGAGTCCCGAGTCCCGAGTCCCGAGTCCCGAG 479  
 QY 426 CCAGCGCGCGGAGCGCTTACAGGTTCTTCCAGTTCGCTTACTTCAAGGCAAAATATCTCCC 485  
 Db 480 CCAGCGCGCGGAGCGCTTACAGGTTCTTCCAGTTCGCTTACTTCAAGGCAAAATATCTCCC 539  
 QY 486 CAAAGCCATCCCGCTGCTCAGACCTTCTTACTGAAGCATCTTGAAGCATCTTGAAGCATCTTGA 545  
 Db 540 CAAAGCCATCCCGCTGCTCAGACCTTCTTACTGAAGCATCTTGAAGCATCTTGAAGCATCTTGA 599  
 QY 546 GAGGAACATGGCATATATAGAGCTTCTTGAAGCTTCTTGAAGCTTCTTGAAGCTTCTTGA 605  
 Db 600 GAGGAACATGGCATATATAGAGCTTCTTGAAGCTTCTTGAAGCTTCTTGAAGCTTCTTGA 659  
 QY 606 AACCAAGTCATATGAAAGCTTGTTCATCCGAGCAGTCCGGGCGATACAAACGCTGAGAACTG 665  
 Db 660 AACCAAGTCATATGAAAGCTTGTTCATCCGAGCAGTCCGGGCGATACAAACGCTGAGAACTG 719  
 QY 666 GAGACATCCATCAGACATGAGCTTGGCGCTTCCGAGCTTCTTCAAGGCTTCTTCAAGGCTTCTTCA 725  
 Db 720 GAGACATCCATCAGACATGAGCTTGGCGCTTCCGAGCTTCTTCAAGGCTTCTTCAAGGCTTCTTCA 779  
 QY 726 GTGCTCGAGCTTCCGAGGCTTCCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTC 785  
 Db 780 GTGCTCGAGCTTCCGAGGCTTCCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTC 839  
 QY 786 CATAGCAGATCATTTATGAGAGTCTTGGAAATGCAAAATACAGTGTGAAGAGAACCTCAC 845  
 Db 840 CATAGCAGATCATTTATGAGAGTCTTGGAAATGCAAAATACAGTGTGAAGAGAACCTCAC 899  
 QY 846 CCCAGTTATAGAGGCTATCCGTTGAGAAATTTG-TGGCTACCATGTATCATTTCTTGC 904  
 Db 900 CCCAGTTATAGAGGCTATCCGTTGAGAAATTTGTTGGCTACCATGTATCATTTCTTGC 959  
 QY 905 AGTTGCTTATATAGTGTGACGCTGAGATGCGAGCTTGGAGCTTGGAGCTTGGAGCTTGGAGCTTGG 964  
 Db 960 AGTTGCTTATATAGTGTGACGCTTGGAGCTTGGAGCTTGGAGCTTGGAGCTTGGAGCTTGGAG 1019  
 QY 965 TGCTCTTTGATCAGAAATGACAAG 987  
 Db 1020 TGCTCTTTKATCAGAAKRAAAG 1042

RESULT 10  
 AK017797  
 LOCUS  
 DEFINITION Mus musculus 8 days embryo whole body cDNA, RIKEN full-length  
 enriched library, clone:570529N23 product:cartilage associated  
 protein, full insert sequence.  
 ACCESSION AK017797  
 VERSION AK017797.1 GI:12857226  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)







```

QY 956 TCAGTATCTCTCTTTGATCAGATGACAAG 987
DB 1032 TMA-STATCTCTBTGATCAGATGACARG 1062

RESULT 13
LOCUS BX386576 339 bp mRNA linear EST 08-MAY-2003
DEFINITION BX386576 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
ACCESSION BX386576 Homo sapiens cDNA clone CS0DJ007YB14 5-PRIME, mRNA sequence.
VERSION BX386576.1 GI:30455453
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 939)
Li, W.B., Gruber, C., Jesses, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2233.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS1AJ002ZF08QP1&cluster=2233.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS1AJ002ZF08QP1.

FEATURES
source
1..939
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ007YB14"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 201 a 291 c 271 g 158 t 18 others
ORIGIN
Query Match 69.7%; Score 838; DB 13; Length 939;
Best Local Similarity 97.7%; Pred. No. 1e-186;
Matches 840; Conservative 17; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATGAGCGCGGGCGCGCGGGCGCGCGCGCTGCTAGCGCTGCTGCGCTGCGCTGCGCG 60
DB 81 ATGAGCGCGGGCGCGCGGGCGCGCGCGCTGCTAGCGCTGCTGCGCTGCGCTGCGCG 140
QY 61 CTGCGCGCGCGGGCGCGCGCGCGCGCGCTGCTAGCGCTGCTGCGCGCGCGCGCGAG 120
DB 141 CTGCGCGCGGGCGCGCGCGCGCGCGCGCTGCTAGCGCTGCTGCGCGCGCGCGAG 200
QY 121 CTGATCGCGCTGCGCTGCGCTGCGCGCGCGCGCGCGCTGCTAGCGCTGCTGCGCG 180
DB 201 CWTGTCGGCYCGAGTCGGCTACCGCGCGCGCGCGCTGCTAGCGCTGCTGCGCGAG 260
QY 181 GCGGAGAGCGCTGCGCTACCTGGAGATCAGCGCTGCGCTGCGCTGCGCTGCGCGAG 240
DB 261 GCGGAGAGCGCTGCGCTACCGCGCGCGCGCGCGCTGCTAGCGCTGCTGCGCGAG 319
QY 241 GAGGCTTGTGCGACCGCACTCAGCGCGCGCGCGCGCGCTGCTAGCGCTGCTGCGCG 300

```

```

DB 320 SAGGCTTCTGCGCACCGCAACTSCAGCGCGCGCGCGCGCGCGCGCGCGCGCTC 379
QY 301 GCAGCTATCCGAGCTGCGCTCTTCGGGGGCGCTGCTGCGCGCGCGCGCGCTCAAG 360
DB 380 GCAGCAWCCGAGCTGCGCTCTWGGGGGCGCTGCTGCGCGCGCGCGCGCTCAAG 439
QY 361 CGTGCACGACGAGGCTGCGCGCTTCCGCCAGTCCCGAGCCCGAGGAGTGTGGCG 420
DB 440 CGTGCACGACGAGGCTGCGCGCTTCCGCCAGTCCCGAGCCCGAGGAGTGTGGCG 499
QY 421 GACTTCCAGCGCGCGAGCGCTTACAGGTTCTCGAGTTGCTTACTTCAAGCAATAAT 480
DB 500 GACTTCCAGCGCGCGAGCGCTTACAGGTTCTCGAGTTGCTTACTTCAAGCAATAAT 559
QY 481 CTCCTCAAGCGCATCGCGCTGCTCACACCTTCTTACTTCAAGCATCTCTGATGACGAATG 540
DB 560 CTCCTCAAGCGCATCGCGCTGCTCACACCTTCTTACTTCAAGCATCTCTGATGACGAATG 619
QY 541 ATGAGAGAGGACATGGCATATATAGAGCGCTGCTGCTGCGGAGGACTACATTAAGAC 600
DB 620 ATGAGAGGACATGGCATATATAGAGCGCTGCTGCTGCGGAGGACTACATTAAGAC 679
QY 601 CTGGAACCAAGTCAATATGAAAGCGCTGCTCATCCGAGCAGTCCGCGCATACACGGTGAG 660
DB 680 CTGGAACCAAGTCAATATGAAAGCGCTGCTCATCCGAGCAGTCCGCGCATACACGGTGAG 739
QY 661 AACTGGAAGACATCCATCACAGACATGAGCGTCCGCGCTTCTTCAAGACCTTT 720
DB 740 AACTGGAAGACATCCATCACAGACATGAGCGTCCGCGCTTCTTCAAGACCTTT 799
QY 721 TACGAGTGTCTGCGAGCGCTGCGAGGTTCCAGGAGAGTCAAGGACTTCAAGGTTCTAC 780
DB 800 TACGAGTGTCTGCGAGCGCTGCGAGGTTCCAGGAGAGTCAAGGACTTCAAGGTTCTAC 859
QY 781 CTTTCCATAGCAGATCAATATGTAGAAGTTCTGGAATGCAGAAATACAGTGTGAAGAGAC 840
DB 860 CTTTCCATAGCAGATCAATATGTAGAAGTTCTGGAATGCAGAAATACAGTGTGAAGAGAC 919
QY 841 CTCACCCCGAGTTATAGGAGG 860
DB 920 CTCACCCCGAGTTATAGGAGG 939

RESULT 14
LOCUS AL558186 1201 bp mRNA linear EST 31-MAY-2003
DEFINITION AL558186 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
ACCESSION AL558186 Homo sapiens cDNA clone CS0DJ001YN15 5-PRIME, mRNA sequence.
KEYWORDS EST.
SOURCE AL558186.2 GI:31279985
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jesses, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2233.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DJ001CG08QP1&cluster=2233.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DJ001CG08QP1.

FEATURES
Location/Qualifiers

```

```
source
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ001Y15"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      279 a 319 c 313 g 228 t
ORIGIN
Query Match      69.3%; Score 834; DB 9; Length 1201;
Best Local Similarity 96.5%; Pred. No. 1.1e-187;
Matches 881; Conservative 20; Mismatches 7; Indels 5; Gaps 5;
QY 1 ATGAGCGCGGCGCGCGCGCGCGCGCGCTGTAGCGCTGCTGCGTGGCGCTGCGTGGCGCTGCGG 60
Db 95 ATGANCCGRRGCGCGGCGCGCGCGCTGTAGCGCTGCTGCGTGGCGCTGCGTGGCGCTGCGG 154
QY 61 CTGCGCGCGCGCGCGCGCGCGCGCAATACAGCGCTACAGCTTCGCGAGCTTCCACGCGGAG 120
Db 155 CTGCGCGCGCGCGCGCGCGCGCGCAATACAGCGCTACAGCTTCGCGAGCTTCCACGCGGAG 214
QY 121 CTGATCGCGCTCGAGTCGGCTACCGGACCGCTGGACAGTACAGCGCGGAGCACTGG 180
Db 215 CTGATCGCGCTCGAGTCGGCTACCGGACCGCTGGACAGTACAGCGCGGAGCACTGG 274
QY 181 GCCAGAGCGCTGCGCTACCTGAGATCAGCTGCGGCTGCACCGCTTGTGCGCGAGCAGC 240
Db 275 GCCAGAGCGCTGCGCTACCTGAGATCAGCTGCGGCTGCACCGCTTGTGCGCGAGCAGC 334
QY 241 GAGCGCTTCTGCCACCGCACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 300
Db 335 GAGCGCTTCTGCCACCGCACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 394
QY 301 GCCAGCTATCCGAGCTGCGCTCTTCGGGGGCGCTGTCGGCGCGCGCGCGCGCGCGCTCAAG 360
Db 395 GCCAGCTATCCGAGCTGCGCTCTTCGGGGGCGCTGTCGGCGCGCGCGCGCGCGCTCAAG 454
QY 361 CGCTGCAAGCAGGCGCTCCAGCGCTCCGCGAGTCCCGCGCGCGCGCGCGCGCGCGCGCTGGCG 420
Db 455 CGCTGCAAGCAGGCGCTCCAGCGCTCCCGCGAGTCCCGCGCGCGCGCGCGCGCGCGCTGGCG 514
QY 421 GACTTCCAGCGCGCGCGCGCGCGCTTCTGCGAGTTCGCTTACTTCAAGGCAATAT 480
Db 515 GACTTCCAGCGCGCGCGCGCGCGCTTCTGCGAGTTCGCTTACTTCAAGGCAATAT 574
QY 481 CTGCGCAAGCGATCGCGCTGCTCAACCTTTCTACTGAGCATCTCTGATGACGAAATG 540
Db 575 CTGCGCAAGCGATCGCGCTGCTCAACCTTTCTACTGAGCATCTCTGATGAGCAATG 634
QY 541 ATGAAGAGGAACATGGCATAT-TATAAGAGCGTG-CCTGGTCCGAGGACTACATTAAAG 598
Db 635 ATGAAGAGGAACATGGCATATATATATATATATATATATATATATATATATATATAT 694
QY 599 ACCTGGAACCAAGTATATGAAGCGCTGTTCATCCGAGCAGTGCAGGCGATACACCGGTG 658
Db 695 ACCTGGAACCAAGTATATGAAGCGCTGTTCATCCGAGCAGTGCAGGCGATACACCGGTG 754
QY 659 AGAAGCTGGAGAACATCATCAGACATGAGTGGCGCTTCCGAGCTTCTTCAAGGCT 718
Db 755 AGAAGCTGGAGAACATCATCAGACATGAGTGGCGCTTCCGAGCTTCTTCAAGGCT 813
QY 719 TTATACGAGTGTCTCGAGCGCTGCGAGGTTTCAGGGAGATCAAGGACTTCAAGGATTTCT 778
Db 814 TTATACGAGTGTCTCGAGCGCTGCGAGGTTTCAGGGAGATCAAGGACTTCAAGGATTTCT 873
QY 779 ACCTTCCATAGCAGATCATTTATGTAGAAGTTCGTGGAATGCAAAATACAGTGTGAAGAGA 838
Db 874 ACBTTCATAGCAGATCATTTATGTAGAAGTTCGTGGAATG-MAATACAGTGTGAAGAGA 932
QY 839 ACCTCACCACCGATTATAGGAGGCTATCCGGTTGAGAAATTTGGCTACCATGTATCAT 898
Db 933 A-CTCACCACCGATTATAGGAGGCTATCCGGTTGAGAAATTTGGCTACCATGTATCAT 991
QY 899 ACTTGCGAGTTTCG 911
Db 992 ACTTGCGAGTTTCG 1004
RESULT 15
AL550966
LOCUS
DEFINITION
AL550966 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0DI065YP05 5-PRIME, mRNA sequence.
ACCESSION
AL550966
VERSION
AL550966.2 GI:31272783
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li W.B., Gruber,C., Jessee,J., and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12888460.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2233.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DI065CH03QPI&cluster=2233.f. Contact :
Peng Liang Email : fliang@lifetech.com URL : Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DI065CH03QPI.
FEATURES
Location/Qualifiers
1..1152
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI065YP05"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      266 a 311 c 285 g 227 t 63 others
ORIGIN
Query Match      68.3%; Score 822.2; DB 9; Length 1152;
Best Local Similarity 96.5%; Pred. No. 6.7e-185;
Matches 898; Conservative 10; Mismatches 16; Indels 7; Gaps 7;
QY 1 ATGAGCGCGGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCGG 60
Db 51 ATGAGCGCGGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCGG 110
QY 61 CTGCGCGCGCGCGCGCGCGCGCGCAATACAGCTTCCGCGAGCTTCCACGCGGAGCAG 120
Db 111 CTGCGCGCGCGCGCGCGCGCGCGCAATACAGCTTCCGCGAGCTTCCACGCGGAGCAG 169
QY 121 CTGATGCGCGCTCGAGTCGGCTACCGGCGCGCGCTGCAAGTACAGCGCGGAGCACTGG 180
Db 170 CTGATGCGCGCTCGAGTCGGCTACCGGCGCGCGCTGCAAGTACAGCGCGGAGCACTGG 229
QY 181 GCCGAGAGCGCTGCGCTACCTGGAGATCAGCGCTGCGCGCTGCACCGCTTGTGCGCGGAGCAGC 240
```

```
Db 230 GCCGAGAGCGTGGGCTACTGGGATCAGCTACGGCTNCACCGCTTGCTCGCGGACAG- 288
Qy 241 GAGGCCCTTCTGCCAAGCGCACTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 300
Db 289 SAGGCCCTTCTGCCAAGCGCACTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCTC 348
Qy 301 GCCAGCTATCCCGAGCTGGCGCTCTTGGGGGCGCTGCTGCGCGCGCGCGCGCGCTCAAG 360
Db 349 GCCAGCWATCCCGAGCTGGCGCTCTTGGGGGCGCTGCTGCGCGCGCGCGCGCTCAAG 408
Qy 361 CGCTGCAAGCAGGCGCTGCGCGCTTCCGCGAGTCCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 409 CGCTGCAA-CAGGGCGTGGCGCGCTTCCGCGAGTCCGCGCGCGCGCGCGCGCGCGCG 467
Qy 421 GACTTCCAGCGCGCGCGCGCTTACAAGTTCCTGCGAGTTCGCTTACTTCAAGGCAAAATAAT 480
Db 468 GACTTCCAGCGCGCGCGCGCTTACAAGTTCCTGCGAGTTCGCTTACTTCAAGGCAAAATAAT 527
Qy 481 CTCGCCAAGCGCATGCCCGCTGCTCAGCTTCTTCTGAGGATCCTGATGACGAAATG 540
Db 528 CTCGCCAAGCGCATGCCCGCTGCTCAGCTTCTTCTGAGGATCCTGATGATGAAATG 587
Qy 541 ATGAGAGGAACATGCGCATATTAT-AAGAGCGCTGCTGCTGCGCGCGCGCGCGCGCTTAAAGA 599
Db 588 ATGAGAGGAACATGCGCATATTATATAAAGCGCTGCTGCTGCGCGCGCGCGCGCGCTTAAAGA 647
Qy 600 CCTGGAACCAAGTCAATATGAAGCGCTTCAATCCGAGCGAGTGGCGCGCATACAACGCTGA 659
Db 648 CCTGGAACCAAGTCAATATGAAGCGCTTCAATCCGAGCGAGTGGCGCGCATACAACGCTGA 707
Qy 660 GAATCGGAGAACATCCATCAGACATGAGGCTGCGCGCTTCCCGACTTCTTCAAAGCGCTT 719
Db 708 GAATCGGAGAACATCCATCAGACATGAGGCTGCGCGCTTCCCGACTTCTTCAAAGCGCTT 767
Qy 720 TTACGAGTCTCGCGAGCTGCGAGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTA 779
Db 768 TTACGAGTCTCGCGAGCTGCGAGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTA 827
Qy 780 CCTTCCATAGCAGATCATATGTAGAGTTCCTGGAATGCAAAATACAGTGTGAAGAGAA 839
Db 828 CCTTCCATAGCAGATCATATGTAG-ARTTCTGGAATGCAAAATACAGTGTGAAGAGAA 886
Qy 840 CTCACCCCGAGTTATAGGAGGCTATCCCGTTGAGAAATTTGGCTACCATGTATCATTA 899
Db 887 -CTCCACCCCGAGTTATAGGAGGCTATCCCGTTGAGAAATTTGGCTACCATGTATCATTA 945
Qy 900 CTTCAGTTTGGCTATTATTAAGTTGAACGAC 930
Db 946 CTTCAGTTTG-CTAATATAGTTGAWSRCC 975
```

Search completed: January 30, 2004, 18:26:30  
Job time : 2528.98 secs



GenCore version 5.1.6  
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 05:50:18 ; Search time 312.764 Seconds  
(without alignments)  
10382.969 Million cell updates/sec

Title: US-09-729-674-1\_COPY\_63\_1265

Perfect score: 1203

Sequence: 1 atggagccggggcgccggg.....tgggaactggagagaccagc 1203

Scoring table:

IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 13Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1202.2	99.9	3871	20	AX60801 Human secreted protein
2	1202.2	99.9	3871	22	AXS59207 Human CDNA encodin
3	1202.2	99.9	3871	24	ABA90876 Human polynucleoti
4	1197.4	99.5	1961	22	AAH14525 Human CDNA sequenc
5	705	58.6	1586	21	AAH31138 Human colon cancer
6	656.6	54.6	818	22	AAH06808 Human CDNA clone
7	465.8	38.7	2347	24	ALH67662 Oesophagus cancer
8	406.4	33.8	408	21	AAH30585 Human colon cancer

9	156.4	13.0	1309	24	ABZ11344 Human polynucleoti
10	156.4	13.0	2152	24	ABZ11345 Human polynucleoti
11	156.4	13.0	2211	24	AAH17573 DNA encoding novel
12	156.4	13.0	2242	24	ABZ11346 Human polynucleoti
13	156.4	13.0	2524	22	AAH14513 Human CDNA sequenc
14	156.4	13.0	2563	22	AAF93800 Human CDNA encodin
15	156.4	13.0	2585	22	AAH14589 Human CDNA sequenc
16	156.4	13.0	2993	22	AAH16288 Human CDNA sequenc
17	154.8	12.9	2600	21	AAC64725 Human tumour suppr
18	154.8	12.9	2829	21	AAC64724 Human tumour suppr
19	152.6	12.7	2577	22	AAH08489 Human secreted pro
20	150.4	12.5	2615	22	AAH08507 Human secreted pro
21	147.4	12.3	2322	21	AAC64727 Mouse tumour suppr
22	147.4	12.3	2416	21	AAC64726 Mouse tumour suppr
23	130.8	10.9	2127	24	AAH17572 DNA encoding novel
24	126	10.5	618	21	AAZ80131 Human colon cancer
25	121.4	10.1	325	22	AAI82020 Human polynucleoti
26	116.2	9.7	1026	24	ABQ26350 Oligonucleotide fo
27	116.2	9.7	1026	24	ABQ26351 Oligonucleotide fo
28	115.8	9.6	1025	24	ABQ27312 Oligonucleotide fo
29	115.8	9.6	1025	24	ABQ27313 Oligonucleotide fo
30	109.2	9.1	2665	25	ACA03965 cDNA downregulated
31	99.2	8.2	652	22	AAF93998 Primer specific fo
32	99.2	8.2	738	22	AAH06612 Human CDNA clone
33	99	8.2	703	22	AAH06792 Human CDNA clone
34	98.6	8.2	774	22	AAH06885 Human CDNA clone
35	98.6	8.2	800	22	AAH06448 Human CDNA clone
36	98.6	8.2	2789	22	AAH14459 Human CDNA sequenc
37	97.8	8.1	2753	24	ABK94910 Human novel polynu
38	97.8	8.1	2753	24	ABK94933 Oligonucleotide fo
39	88	7.3	1025	24	ABQ27310 Oligonucleotide fo
40	88	7.3	1025	24	ABQ27311 Oligonucleotide fo
41	88	7.3	1026	24	ABQ26352 Oligonucleotide fo
42	88	7.3	1026	24	ABQ26353 Oligonucleotide fo
43	87	7.2	349	21	AAH18376 Lung cancer associ
44	73.2	6.1	550	25	ACA57060 Human adipocyte se
45	68.2	5.7	3897	24	ABL60615 Maize SBMu200 geno

#### ALIGNMENTS

#### RESULT 1

AX60801

ID AAX60801 standard; DNA; 3871 BP.

XX

AC AAX60801;

XX

DT 09-AUG-1999 (first entry)

XX

DE Human secreted protein encoding DNA (clone bd306-7).

XX

KW Secreted protein; kidney; lung; brain; blood; testis; bone marrow;

KW nutritional activity; cytokine; cell proliferation; immune stimulation;

KW hematopoiesis regulation; tissue growth; thrombolytic; gene therapy;

KW anti-inflammatory; tumour invasion; ss.

XX Homo sapiens.

OS

PN WO9926961-A1.

XX

PD 03-JUN-1999.

XX

PF 24-NOV-1998; 98WO-US25149.

XX

FR 23-NOV-1998; 98US-0197886.

PR 26-NOV-1997; 97US-0066804.

XX

PA (GEM ) GENETICS INST INC.

XX

PI Agostino MJ, Clark HF, Collins-Racie LA, Evans C;

PI Rechel K, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PI Steininger RJ, Treacy M, Wong GG;



PD 11-OCT-2001.  
XX 22-MAR-2001; 2001WO-US09369.  
PF 30-MAR-2000; 2000US-0539330.  
PR 04-DEC-2000; 2000US-0729674.  
XX (GEMV) GENETICS INST INC.  
PA  
XX  
PI Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;  
PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;  
PI Clark H, Fichtel K, Merberg D;  
XX  
DR WPI; 2001-639363/73.  
DR P-PSDB; AAU38989.  
XX  
PT Secreted human proteins, useful as vaccine for treating various  
PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and  
PT nervous system disorders (e.g. stroke) -  
XX  
PS Claim 1; Page 456-457; 619pp; English.  
XX  
CC The invention relates to novel human secreted proteins, the nucleic  
CC acids encoding them. The protein may exhibit cytokine, cell proliferation  
CC or cell differentiation activity or may induce production of other  
CC cytokines in certain cell populations and may exhibit immune stimulating  
CC or immune suppressing activity, which is useful for the treatment of  
CC various immune deficiencies and disorders e.g. severe combined  
CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,  
CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary  
CC inflammation. The proteins are also useful in the treatment of diseases  
CC and disorders including tissue, skin and organ transplantation and in  
CC graft-versus-host diseases (GVHD), in the induction of tumour immunity,  
CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,  
CC in the treatment of burns, incisions and ulcers; as well as in treatment  
CC of periodontal disease, osteoporosis or osteoarthritis, mediated by  
CC inflammatory processes, diseases of the peripheral nervous system,  
CC Alzheimer's, Parkinson's disease, Huntington's disease,  
CC amyotrophic lateral sclerosis, and Shy-Drager syndrome, infections,  
CC infarction of cardiac and central nervous system vessel e.g. stroke,  
CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The  
CC protein, having activin- or inhibin-related activities is useful as a  
CC contraceptive based on the ability of inhibin to decrease fertility in  
CC female mammals and decrease spermatogenesis in male mammals. The  
CC proteins and nucleic acids are also useful as food supplements. The  
CC present sequence encodes a secreted protein of the invention.  
XX  
SQ Sequence 3871 BP; 965 A; 913 C; 1033 G; 958 T; 2 other;

Query Match 99.9%; Score 1202.2; DB 22; Length 3871;  
Best Local Similarity 100.0%; Pred. No. 1.7e-253;  
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCGGGGCGCGGGGGCGCGCGCGCTGCTAGCGTGTGCGTGGCGCTCGCG 60  
DB 63 ATGAGCGGGGCGCGGGGGCGCGCGCGCTGCTAGCGTGTGCGTGGCGCTCGCG 122  
QY 61 CTGCGCGCGGGCGCGCGCGCGCGCGCGCTGCTAGCGTGTGCGTGGCGCGAG 120  
DB 123 CTGCGCGCGGGCGCGCGCGCGCGCGCGCTGCTAGCGTGTGCGTGGCGCGAG 182  
QY 121 CTGATGCGGCTCGAGTGGCGCTACCGGCGCGCGCGCTGCTAGCGTGTGCGTGG 180  
DB 183 CTGATGCGGCTCGAGTGGCGCTACCGGCGCGCGCGCTGCTAGCGTGTGCGTGG 242  
QY 181 GCGAGAGCGTGGCTACCTGGAGATCAGCTCGGCTGCGCGCTGCTGGCGCGAGC 240  
DB 243 GCGAGAGCGTGGCTACCTGGAGATCAGCTCGGCTGCGCGCTGCTGGCGCGAGC 302  
QY 241 GAGCGCTTCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 300  
DB 303 GAGCGCTTCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 362

## RESULT 3

ABA90876  
ID ABA90876 standard; cDNA; 3871 BP.  
XX  
AC ABA90876;  
XX  
DT 14-FEB-2002 (first entry)  
XX

QY 301 GCCAGCTATCCGAGCTGCGCTCTTTCGGGGGCGCTGCTGCGGCGCGCGCACTGCGCTCAAG 360  
DB 363 GCCAGCTATCCGAGCTGCGCTCTTTCGGGGGCGCTGCTGCGGCGCGCGCACTGCGCTCAAG 422  
QY 361 CGCTGCAAGCAGGCGCTGCGCGCTTCCGCGAGTCCGCGCGCGCGCGCGCGAGGTGCTGGCG 420  
DB 423 CGCTGCAAGCAGGCGCTGCGCGCTTCCGCGAGTCCGCGCGCGCGCGCGCGAGGTGCTGGCG 482  
QY 421 GACTTCCAGCGCGCGCGCGCTTACAGTTCCTGCGAGTTCGCTTACTTCAAGGCAATAAT 480  
DB 483 GACTTCCAGCGCGCGCGCGCTTACAGTTCCTGCGAGTTCGCTTACTTCAAGGCAATAAT 542  
QY 481 CTCCCAAGAGCCATCGCGCTGCTCACCTTCTACTGAAGCATCTCTGATGACGAAATG 540  
DB 543 CTCCCAAGAGCCATCGCGCTGCTCACCTTCTACTGAAGCATCTCTGATGACGAAATG 602  
QY 541 ATGAAGAGCAACATGGCATATTATTAAGAGCCCTGCTGCGCGAGGAGTACATTAAGAGC 600  
DB 603 ATGAAGAGCAACATGGCATATTATTAAGAGCCCTGCTGCGCGAGGAGTACATTAAGAGC 662  
QY 601 CTGGAACCAAGTCAATGAAAGCCCTGTTTATCCGAGCAGTTCGGGCGCATACACGCTGAG 660  
DB 663 CTGGAACCAAGTCAATGAAAGCCCTGTTTATCCGAGCAGTTCGGGCGCATACACGCTGAG 722  
QY 661 AACTGGAGAACATCCATCACAGCATGAGCTGGCGCTTCCGCGACTTCTTCAAGGCTTT 720  
DB 723 AACTGGAGAACATCCATCACAGCATGAGCTGGCGCTTCCGCGACTTCTTCAAGGCTTT 782  
QY 721 TAGAGTGTCTCGCAGCCCTCGAGGGTTCGAGGAGATCAAGGACTTCAAGGATTTCTAC 780  
DB 783 TAGAGTGTCTCGCAGCCCTCGAGGGTTCGAGGAGATCAAGGACTTCAAGGATTTCTAC 842  
QY 781 CTTTCCATAGCAGATCATTTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAGAAC 840  
DB 843 CTTTCCATAGCAGATCATTTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAGAAC 902  
QY 841 CTCACCCAGTTATAGGAGGCTATCCGTTGAGAAATTTGCTGCTACCATGTATCATTTAC 900  
DB 903 CTCACCCAGTTATAGGAGGCTATCCGTTGAGAAATTTGCTGCTACCATGTATCATTTAC 962  
QY 901 TTGCGATTTGCTTATTAAGTTGAACCGCTGGAAGATGAGCGCGCTGCTGCGAGTCAGC 960  
DB 963 TTGCGATTTGCTTATTAAGTTGAACCGCTGGAAGATGAGCGCGCTGCTGCGAGTCAGC 1022  
QY 961 TATCTGCTTTTGTATCAGAAATGACAGTTCATGCGAGAGACCTTGTGTATTTACAGTAC 1020  
DB 1023 TATCTGCTTTTGTATCAGAAATGACAGTTCATGCGAGAGACCTTGTGTATTTACAGTAC 1082  
QY 1021 CACAGGACACTTGGGGGCTCTCGGATGAGCACTTCCAGCGCGCGAGACCTTGAAGCAGTTTAC 1080  
DB 1083 CACAGGACACTTGGGGGCTCTCGGATGAGCACTTCCAGCGCGCGAGACCTTGAAGCAGTTTAC 1142  
QY 1081 TTCTTTATGTCACACACTCCAGAGGAGCTGTATGACTTTGTAAGGAAATATTAATG 1140  
DB 1143 TTCTTTATGTCACACACTCCAGAGGAGCTGTATGACTTTGTAAGGAAATATTAATG 1202  
QY 1141 GATGATGATGAGGAGAGAGTTGTGGAATATGATGATGATGATGATGATGATGATGATGATG 1200  
DB 1203 GATGATGATGAGGAGAGAGTTGTGGAATATGATGATGATGATGATGATGATGATGATG 1262  
QY 1201 AGC 1203  
DB 1263 AGC 1265

DE Human polynucleotide SEQ ID NO 1.  
XX  
KW Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;  
KW immune disorder; bacterial infection; fungal infection; cancer; tumour;  
KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibitor;  
KW osteoarthritis; osteoarthritis; nervous system disorder; neuropathy;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;  
KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnerability;  
KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;  
KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;  
KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.  
XX  
OS Homo sapiens.  
XX  
US2001039335-A1.  
XX  
PD 08-NOV-2001.  
XX  
XX 04-DEC-2000; 2000US-0729674.  
XX  
PR 10-APR-1997; 97US-126425P.  
PR 04-DEC-1997; 97US-067454P.  
PR 20-DEC-1997; 97US-068379P.  
PR 02-JAN-1998; 98US-070346P.  
PR 07-JAN-1998; 98US-070643P.  
PR 08-JAN-1998; 98US-070755P.  
PR 13-JAN-1998; 98US-071304P.  
PR 22-JAN-1998; 98US-072113P.  
PR 30-JAN-1998; 98US-073095P.  
PR 18-FEB-1998; 98US-075038P.  
PR 30-MAR-2000; 2000US-0539330.  
PR 23-NOV-1998; 98US-0197886.  
XX  
(JACO/) JACOBS K.  
PA (MCCO/) MCCOY J M.  
PA (LAVA/) LAVALLIE E R.  
PA (COLL/) COLLINS-RACIE L A.  
PA (EVAN/) EVANS C.  
PA (MERB/) MERBERG D.  
PA (TREA/) TREACY M.  
PA (AGOS/) AGOSTINO M J.  
PA (STEI/) STEININGER R J.  
PA (SPAU/) SPAULDING V.  
PA (WONG/) WONG G G.  
PA (CLAR/) CLARK H.  
PA (FECH/) FECHTEL K.  
XX  
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;  
PI Wong GG, Clark H, Fechtel K;  
XX  
DR WPI: 2002-040725/05.  
DR P-PSDB; ABB55698.  
XX  
PT New secreted proteins and encoding polynucleotides, useful in gene  
PT therapies, particularly for preventing or treating autoimmune  
PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,  
PT stroke or inflammations  
XX  
PS Claim 1; Page 175-177; 349pp; English.  
XX  
CC The invention relates to isolated polynucleotides (ABA90876-ABA90568 and  
CC ABA90980) and encoded proteins (ABB55698-ABB55900), especially  
CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and  
CC proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in  
CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1  
CC are deposited with the American Type Culture Collection (ATCC) with  
CC accession number 98599. The polynucleotides and encoded polypeptides have  
CC cytosolic, anti-inflammatory, immunomodulator, vulnerability,  
CC neuroprotective, activin, inhibitor, chemotactic, haemostatic, thrombolytic  
CC and anti-inflammatory activity and acting as cytokine modulators,  
CC haematopoiesis regulators, tissue growth modulators and/or cadherin  
CC suppressors. The polypeptides and polynucleotides are useful in gene

CC therapies, particularly for preventing, treating or ameliorating any of  
CC the following diseases: immune deficiency and disorders; e.g. bacterial  
CC or fungal infections, autoimmune disorders, cancer, systemic lupus  
CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell  
CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
CC osteoarthritis; central and peripheral nervous system diseases or  
CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;  
CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis  
CC or systemic inflammatory response syndrome; ischaemia-reperfusion  
CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or  
CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus  
CC foliaceus.  
XX  
SQ Sequence 3871 BP; 965 A; 913 C; 1033 G; 958 T; 2 other;  
Query Match 99.9%; Score 1202.2; DB 24; Length 3871;  
Best Local Similarity 100.0%; Pred. NO. 1.7e-253;  
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGAGCCGGGGCGCCGGGGGGCGGGCGCTCTAGCGCTGTGTGCGTGGCTCGCG 60  
Db |||  
63 ATGGAGCCGGGGCGCCGGGGGGCGGGCGCTCTAGCGCTGTGTGCGTGGCTCGCG 122  
QY 61 CTGCGCGCCGGGGCGCCGGGGGGCGGGCGCTCTAGCGCTGTGTGCGTGGCTCGCG 120  
Db |||  
123 CTGCGCGCCGGGGCGCCGGGGGGCGGGCGCTCTAGCGCTGTGTGCGTGGCTCGCG 182  
QY 121 CTGATGCGCGCTCGAGTGGCGCTACCGGCGACCGCTCGCAAGTACAGCGCGAGCACTGG 180  
Db |||  
183 CTGATGCGCGCTCGAGTGGCGCTACCGGCGACCGCTCGCAAGTACAGCGCGAGCACTGG 242  
QY 181 GCGGAGAGCSTKGCTACCTGGAGATCAGCTCGGCTGCGACCGCTTGTGTGCGCGACAGC 240  
Db |||  
243 GCGGAGAGCSTKGCTACCTGGAGATCAGCTCGGCTGCGACCGCTTGTGTGCGCGACAGC 302  
QY 241 GAGGCGCTTCTGCCACCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 300  
Db |||  
303 GAGGCGCTTCTGCCACCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 362  
QY 301 GCAGCTATCCGAGTGGCGCTCTTTCGGGGGCGCTGCTCGCGGGCGCGCACTGCTCAAG 360  
Db |||  
363 GCAGCTATCCGAGTGGCGCTCTTTCGGGGGCGCTGCTCGCGGGCGCGCACTGCTCAAG 422  
QY 361 CGTGAAGCAGCGCGCTGCGCGCGCTTCCGCGAGTCCGCGCGCGCGCGCGCGCGCGCGCG 420  
Db |||  
423 CGTGAAGCAGCGCGCTGCGCGCGCTTCCGCGAGTCCGCGCGCGCGCGCGCGCGCGCGCG 482  
QY 421 GACTTCCAGCGCGCGCGCGCGCTTCCGCGAGTCCGCGAGTCCGCGCGCGCGCGCGCGCGCG 480  
Db |||  
483 GACTTCCAGCGCGCGCGCGCGCTTCCGCGAGTCCGCGAGTCCGCGCGCGCGCGCGCGCGCG 542  
QY 481 CTCGCCAAGAGCCATCGCGCTGCTCACACCTTTCTACTGAAGCATCTGATGACGAATG 540  
Db |||  
543 CTCGCCAAGAGCCATCGCGCTGCTCACACCTTTCTACTGAAGCATCTGATGACGAATG 602  
QY 541 ATGAAGAGGAACATGCGCATATTATGAAGCGCTGCTGCGCGAGGACTACATTAAGAGAC 600  
Db |||  
603 ATGAAGAGGAACATGCGCATATTATGAAGCGCTGCTGCGCGAGGACTACATTAAGAGAC 662  
QY 601 CTGGAACCAAGTCATATGAAGCGCTTTCATCCGAGCAGTGGCGGCATACACCGGTGAG 660  
Db |||  
663 CTGGAACCAAGTCATATGAAGCGCTTTCATCCGAGCAGTGGCGGCATACACCGGTGAG 722  
QY 661 AACTGGAGAACATCCATCACAGACATGGAGTGGCGCTTCCGCGCTTCTTCAAGAGCCCTT 720  
Db |||  
723 AACTGGAGAACATCCATCACAGACATGGAGTGGCGCTTCCGCGCTTCTTCAAGAGCCCTT 782  
QY 721 TACGAGTGTCTCGAGCCTGGAGGGTTCAGGGAGATCAAGGACTTCAAGGATTTCTAC 780  
Db |||  
783 TACGAGTGTCTCGAGCCTGGAGGGTTCAGGGAGATCAAGGACTTCAAGGATTTCTAC 842  
QY 781 CTTCCATAGCAGATCATATTATGTAAGTTCGTGAATTCGAAATCAAAATACAGTGTGAGAGAAC 840



Db 700 AACTGGAGAACATCATCATCAGACATGAGCTGGCCCTTCCGACTTCTTCAAGGCTTT 759  
QY 721 TACAGTGTCTCGCAGCTCGAGGGTTCCAGGAGATCAAGCACTTCAAGGATTTCTAC 780  
Db 760 TACAGTGTCTCGCAGCTCGAGGGTTCCAGGAGATCAAGCACTTCAAGGATTTCTAC 819  
QY 781 CTCTCCATACAGATCATTAATAGTAGAGTTCTCGAATGCAAAATACAGTGTGAAGAAC 840  
Db 820 CTCTCCATACAGATCATTAATAGTAGAGTTCTCGAATGCAAAATACAGTGTGAAGAAC 879  
QY 841 CTCACCCAGTTATAGAGGCTATCCGTTGAGAAATTTGTGCTACCATGTATCATAC 900  
Db 880 CTCACCCAGTTATAGAGGCTATCCGTTGAGAAATTTGTGCTACCATGTATCATAC 939  
QY 901 TTGAGTTTCCCTATTATAGTTGAACGCTTGAAGAAATGCAAGCTTGTGCTGAGT 960  
Db 940 TTGAGTTTCCCTATTATAGTTGAACGCTTGAAGAAATGCAAGCTTGTGCTGAGT 999  
QY 961 TATCTGCTCTTTGATCAGATGACAGGTCATGACAGAACTGTGTTATACAGTAC 1020  
Db 1000 TATCTGCTCTTTGATCAGATGACAGGTCATGACAGAACTGTGTTATACAGTAC 1059  
QY 1021 CACAGGACACTTGGGGCTCTCGATGAGCACTTCCAGCCAGACCTGGAAGCTTCAG 1080  
Db 1060 CACAGGACACTTGGGGCTCTCGATGAGCACTTCCAGCCAGACCTGGAAGCTTCAG 1119  
QY 1081 TTCTTTAATGTGACCACTCCAGAGGAGCTGTATGATTTGCTTAAGAAATATATG 1140  
Db 1120 TTCTTTAATGTGACCACTCCAGAGGAGCTGTATGATTTGCTTAAGAAATATATG 1179  
QY 1141 GATGATGATGAGGAGAACTTGTGGAATATGTGGATGACCTCTTGGAACTGGAGAGACC 1200  
Db 1180 GATGATGATGAGGAGAACTTGTGGAATATGTGGATGACCTCTTGGAACTGGAGAGACC 1239  
QY 1201 AGC 1203  
Db 1240 AGC 1242

## RESULT 5

AAH31138

ID AAH31138 standard; cDNA; 1586 BP.

AC AAH31138;  
XX

27-JUL-2001 (first entry)

Human colon cancer cell line Km12L4-A cDNA library derived seq#1072.

Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;  
detection; colon cancer cell line Km12L4-A; ss.

XX Homo sapiens.

XX WO200018916-A2.

XX 06-APR-2000.

XX 23-SEP-1999; 98WO-US22226.

XX 28-SEP-1998; 98US-0102161.

XX 28-SEP-1998; 98US-0102180.

XX 29-SEP-1998; 98US-0102380.

XX 08-OCT-1998; 98US-0103815.

XX 27-OCT-1998; 98US-0105877.

XX (CHIR ) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
XX WPI; 2000-293155/25.  
XX Polynucleotide library comprising 1079 defined sequences, useful in  
PT the form of an array to detect cancer or susceptibility to cancer -  
XX Claim 1; Page 498-499; 502pp; English.  
XX The present invention describes a library of polynucleotides comprising  
CC 1079 nucleotide sequences (Given in AAH30067 to AAH31145). Also described  
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to  
CC one of the 1079 sequences; (2) a recombinant host cell containing (1);  
CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that  
CC specifically binds to (II); (5) a vector comprising (1); and (6) a method  
CC of detecting differentially expressed genes correlated with a cancerous  
CC state of a mammalian cell comprising detecting a gene product encoded by  
CC 65 of the 1079 sequences given in the specification. The polynucleotides  
CC are used to monitor patients having (or susceptible) to cancer to detect  
CC potentially malignant events at a molecular level before they are  
CC detectable at a gross morphological level. The polynucleotides are also  
CC useful for monitoring the efficacy of various therapies and preventive  
CC interventions. Polynucleotide probes based on the disclosed sequences  
CC are useful for chromosome mapping and detection of transcription levels.  
CC The 1079 polynucleotide sequences were derived from a human colon cancer  
CC cell line Km12L4-A cDNA library.

SQ Sequence 1586 BP; 381 A; 406 C; 340 G; 453 T; 6 other;

Query Match 58.6%; Score 705; DB 21; Length 1586;

Best Local Similarity 99.3%; Pred. No. 9.1e-145;

Matches 702; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 497 CCCTGCTCACACCTTTCTACTGAAGCATCCTGATGAGAAATGATGAAGAGAAATCG 556  
Db 1 CCGCTGCTCACACCTTTCTACTGAAGCATCCTGATGAGAAATGATGAAGAGAAATCG 60  
QY 557 CATATTATAGAGCTGCTGCTGCGGAGACTACATTAAAGACCTGGAAACCAAGTCAT 616  
Db 61 CATATTATAGAGCTGCTGCTGCGGAGACTACATTAAAGACCTGGAAACCAAGTCAT 120  
QY 617 ATGAAAGCCTGTTTCATCCGAGCAGTGGCGGCATACAAAGCTGAGAACTGGAAACATCCA 676  
Db 121 ATGAAAGCCTGTTTCATCCGAGCAGTGGCGGCATACAAAGCTGAGAACTGGAAACATCCA 180  
QY 677 TCACAGACATGAGCTGGCCCTTCCGACTTCTTCAAAGCCTTTTACGAGTCTTCGAG 736  
Db 181 TCACAGACATGAGCTGGCCCTTCCGACTTCTTCAAAGCCTTTTACGAGTCTTCGAG 240  
QY 737 CCTGCGAGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAGATC 796  
Db 241 CCTGCGAGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAGATC 300  
QY 797 ATTATGTAGAAGTCTTGAATGCAAAATACAGTGTGAAGAGAACTCCACCCAGTTATAG 856  
Db 301 ATTATGTAGAAGTCTTGAATGCAAAATACAGTGTGAAGAGAACTCCACCCAGTTATAG 360  
QY 857 GAGGCTATCCGCTTGAGAAATTTGCGCTACCATGTATCATTTAGTTCAGTTCGCTATT 916  
Db 361 GAGGCTATCCGCTTGAGAAATTTGCGCTACCATGTATCATTTAGTTCAGTTCGCTATT 420  
QY 917 ATAAATTGAACACTGGAAGATGCGAGCCCTGCGAGTCAGCTATCTGCTCTTGTATC 976  
Db 421 ATAAATTGAACACTGGAAGATGCGAGCCCTGCGAGTCAGCTATCTGCTCTTGTATC 480  
QY 977 AGAATGACAAGTCTATGCGAGAGACCTGGTGTATTACCATGACAGGAGACCTTGGG 1036  
Db 481 AGAATGACAAGTCTATGCGAGAGACCTGGTGTATTACCATGACAGGAGACCTTGGG 540  
QY 1037 GCCTTCGAGTACGACCTTCCAGGCCCGAGACCTGGAAGAGTTCAGTTCCTTTAATGTACCA 1096  
Db 541 GCCTTCGAGTACGACCTTCCAGGCCCGAGACCTGGAAGAGTTCAGTTCCTTTAATGTACCA 600









Db 841 CCAATTGACCCCAATGTGGGTGGCTACTTCTGTGACCAAGTTCGTGGCCACCATGTACC 900  
 QY 896 ATTACTTCAGCTTGGCTATTATTAAGTTGAAGACCTGAAGATGACAGCCCTCTGTGCAG 955  
 Db 901 ACTACTTCAGCTTGGCTATTATTAAGTTGAAGATGACAGCCCTCTGTGCAG 960  
 QY 956 TCAGCTATCTGCTCTTTGATGACAAATGACAAAGTATGACAGCAAGCTGCTGTATTACC 1015  
 Db 961 CCAGCTACATGCTCTTCGACCCCAAGGACAGGCTCATGACAGCAAGCTGCTGTATTACC 1020  
 QY 1016 AGTACACAGGACACACTTGGGCTCTCTCGATGAGCACTTCAGCCCAAGCTGAAGCAG 1075  
 Db 1021 GGTTCACCGGCTCTCTGGGCTCTGGAAGAGGAGGACTTCAGCCCGGAGGAGGCA 1080  
 QY 1076 TTCACTTCTTTAATGTGACCAACACCTCCAGAGGAGCTGTATGACTTTCTTAAGGAATA 1135  
 Db 1081 TGCTCTACCAACACAGCCCGGAGCTGCGGAGCTGCTGAGTTCAACCACATGTACC 1140  
 QY 1136 TAAATGATGATGATGAGGAGAGTTGTGGA 1166  
 Db 1141 TGCAGTCAGATGATGAGATGGAGCTGGAGGA 1171

## RESULT 8

AAH30585

ID AAH30585 standard; cDNA; 408 BP.

 XX  
 AC AAH30585;

27-JUL-2001 (first entry)

Human colon cancer cell line Kml2L4-A cDNA library derived sequence #519.

Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;

detection; colon cancer cell line Kml2L4-A; ss.

OS Homo sapiens.

PN W0200018916-A2.

06-APR-2000.

23-SEP-1999; 99WO-US22226.

28-SEP-1998; 98US-0102161.

28-SEP-1998; 98US-0102180.

29-SEP-1998; 98US-0102380.

08-OCT-1998; 98US-0103815.

27-OCT-1998; 98US-0105877.

(CHIR) CHIRON CORP.

(HYSE-) HYSEQ INC.

 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
 PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
 XX WPI; 2000-293155/25.

DR WPI; 2000-293155/25.

 XX Polynucleotide library comprising 1079 defined sequences, useful in  
 PT the form of an array to detect cancer or susceptibility to cancer -

PS Claim 1; Page 333; 502pp; English.

 XX The present invention describes a library of polynucleotides comprising  
 CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described  
 CC are: (1) an isolated polynucleotide (I) having at least 90% identity to  
 CC one of the 1079 sequences; (2) a recombinant host cell containing (1);  
 CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that  
 CC specifically binds to (II); (5) a vector comprising (I); and (6) a method  
 CC of detecting differentially expressed genes correlated with a cancerous  
 CC state of a mammalian cell comprising detecting a gene product encoded by

CC 65 of the 1079 sequences given in the specification. The polynucleotides  
 CC are used to monitor patients having (or susceptible) to cancer to detect  
 CC potentially malignant events at a molecular level before they are  
 CC detectable at a gross morphological level. The polynucleotides are also  
 CC useful for monitoring the efficacy of various therapies and preventive  
 CC interventions. Polynucleotide probes based on the disclosed sequences  
 CC are useful for chromosome mapping and detection of transcription levels.  
 CC The 1079 polynucleotide sequences were derived from a human colon cancer  
 CC cell line Kml2L4-A cDNA library.

SQ Sequence 408 BP; 117 A; 96 C; 98 G; 97 T; 0 other;

## Query Match

Best Local Similarity 33.8%; Score 406.4; DB 21; Length 408;

Matches 407; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 497 CCCTGTCTCACACCTTTCTACTGAAGCATCCTGATGACGAATGATGAAGAGGAACATG 556  
 Db 1 CCCTGTCTCACACCTTTCTACTGAAGCATCCTGATGACGAATGATGAAGAGGAACATG 60  
 QY 557 CATATTATTAAGAGCTGCTGTGTCGAGGACTACATTAAAGACCTGGAAACCAAGTCAT 616  
 Db 61 CATATTATTAAGAGCTGCTGTGTCGAGGACTACATTAAAGACCTGGAAACCAAGTCAT 120  
 QY 617 ATGAAAGCCTGTTTCATCCGAGCAGTCCGGGCATACAAACGTCGAGACTGGAGAACATCCA 676  
 Db 121 ATGAAAGCCTGTTTCATCCGAGCAGTCCGGGCATACAAACGTCGAGACTGGAGAACATCCA 180  
 QY 677 TCACAGACATGAGGTGCGCTTCCCGACTTCTTCAAAGCCTTTACGAGTGTCTCGCAG 736  
 Db 181 TCACAGACATGAGGTGCGCTTCCCGACTTCTTCAAAGCCTTTACGAGTGTCTCGCAG 240  
 QY 737 CTGCGAGAGTTCACGGGAGATCAAGGACTTCAAGGATTTACCTTTCCATAGCAGATC 796  
 Db 241 CTGCGAGAGTTCACGGGAGATCAAGGACTTCAAGGATTTACCTTTCCATAGCAGATC 300  
 QY 797 ATTATGTAGAGTTCCTGCAATGCAAAATACAGTGTGAAGAGAACCTCACCCTGATTATAG 856  
 Db 301 ATTATGTAGAGTTCCTGCAATGCAAAATACAGTGTGAAGAGAACCTCACCCTGATTATAG 360  
 QY 857 GAGGCTATCCGTTGAGAAATTTGTGGCTACCATATCATCTATCTTGC 904  
 Db 361 GAGGCTATCCGTTGAGAAATTTGTGGCTACCATATCTATCTTGTGC 408

## RESULT 9

ABZ11344

ID ABZ11344 standard; cDNA; 1309 BP.

XX ABZ11344;

 XX  
 DT 20-JAN-2003 (first entry)

Human polynucleotide SEQ ID NO 226.

 DE Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 XX cell-proliferative disorder; neurodegenerative disease; bacterial;  
 XX Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 KW arthritis; cyclostatic; immunomodulator; nootropic; neuroprotective;  
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
 KW antiarthritic; gene; ss.

OS Homo sapiens.

 XX  
 PN W0200270539-A2.

 XX  
 PD 12-SEP-2002.

 XX  
 PF 05-MAR-2002; 2002WO-US05085.

 XX  
 PR 05-MAR-2001; 2001US-0799451.

XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX  
DR WPI: 2002-759812/82.  
DR P-PSDB; ABP69127.  
XX  
XX New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
PT platelet or coagulation disorders  
XX  
PS Claim 1; SEQ ID NO 226; 1012pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences  
CC (ABZ11119-ABZ12066) or their mature protein coding portion, active domain  
CC coding protein or complementary sequences. The polynucleotides are useful  
CC for identifying expressed genes or for physical mapping of human genome.  
CC The encoded polypeptides (ABP6902-ABP6949) are useful as molecular  
CC weight markers, as a food supplement, for generating antibodies, in  
CC medical imaging, screening and diagnostic assays and for treating  
CC cell-proliferative disorders (cancer), neurodegenerative diseases  
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
CC disorders, platelet or coagulation disorders, wound, burns, incision,  
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
CC parasitic), arthritis, etc.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1309 BP; 287 A; 384 C; 365 G; 273 T; 0 other;  
  
Query Match 13.0%; Score 156.4; DB 24; Length 1309;  
Best Local Similarity 53.6%; Pred. NO. 1.1e-24;  
Matches 437; Conservative 0; Mismatches 346; Indels 33; Gaps 4;  
  
QY 222 CCGCTTGTGTCGCGCAGCGAGGCTTCTGCGACCGCACTGACGCGCGCGCGCGCGCC 281  
DB 256 CGCACCCAGTGTGCGCGCGACTTCCGTGGAGCTGGACCCCGACTGTCTCCCGACGCC 315  
  
QY 282 CGAGCCCGCGCGCGCTTCCGAGCTATCCCGAGCTGCGCTTTCGGGGCGCTCTGTCGCG 341  
DB 316 GCGCCAGCGCTCGGCGCGCGCGCGCTTCCGAGCTGCGCTTCTTCGGGGCGCTCTGCGG 375  
  
QY 342 CCGCGCGCACTGCTCAGCGCTGCAAGCGCGCGCTGCGAGCTTCCGCGAGTCCCGAGCC 401  
DB 376 TCGGCTGCTGCTGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426  
  
QY 402 CAGCGCGAGTGTGCGCGGCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 461  
DB 427 CAGCGAGAGATGAGCTGGAGTTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486  
  
QY 462 TTACTTCAAGCAATATCTCCCAAGCATCGCGCTGCTCAGACTTCTTCTGAGTTCGAA 521  
DB 487 CTACTTCAAGATCAAGATGGAGAGCTGTTGCTGCGAGCACACACTTCTTCTGCTGGG 546  
  
QY 522 GCATCTCTGATGACGAATATGATGAAGAGCAATGGCATATATTAAGAGCTGCTGCTGGT- 580  
DB 547 CAATCTTGAGCACATGGAATGACAGAGAGCTAGACTATTAACAACCAATGCTGAGT 606  
  
QY 581 --CCGAGCATACATTAAGCACTGGAAACCAAGTATATGAAGAGCTGTTATCCGAGC 638  
DB 607 GAAGAGGCGCGACTTCAAGATCTTGGAGCTCAACCCCATATGCAAGAAATTCGACTGGG 666  
  
QY 639 AGTGGCGGCATACAAACGCTGAGAACTGGAGAAATCCATCCATCAGACATGAGCTGCGCT 698  
DB 667 AGTGGGACTTCTCAGAGGAGACGCCACAGGAGAGCTGTGCCCCACTTAGAGCGCGCT 726

699 TCCGAGCTTCTTCAAAGCCCTTTACGAGTGTCTCGCAGCTCGAGGGTTCCAGGAGAT 758  
727 GCAAGATATCTTGTGCGCTATGAGAGTGCCTGCGCTTGCAGAGGCGCTTATGACTA 786  
759 CAGGAGCTTCAA-----GGATTTCTACCTTTCCATAGCAGATCATTA 800  
787 CGATGGCTACAACTACCTTGAGTACAAACGCTGACCTTCCAGGCGCATCACAGATCATTA 846  
801 TGTAGAGATTTCTGGAATGCAAAATACAGTGTGAAGAGAACCT--CACCCAGTTATAGG 857  
847 CATCCAGTCTCTCACTGTAAGCAGAACTGTGTACGAGAGCTTGTCCACCAAGTCG 906  
858 AGCTATTCGGTTGAGAAATTTGTGCTACCATGTATCATTTCCAGTTGCTTATTA 917  
907 AGAAGAGCCCTTTGAAGACTTCTCCATCGCAATTAATATCTCAGTTTGCCTACTA 966  
918 TAAGTTGAACGACCTGAAGATGCAAGCCCTGTCAGTGCAGTATCTCTTTGATCA 977  
967 TAACATGGGAATATACAGAGCTTTGAATGTGCCAAGACTATCTTCTTCTTCTCC 1026  
978 GAATGCAAGGTCACTCAGCAGAACCTGCTGTATTA 1013  
1027 CAATGACGAGGTGATGAACCAAAATTTGCTTATTA 1062

RESULT 10  
ABZ11345  
ID ABZ11345 standard; cDNA; 2152 BP.  
XX ABZ11345;  
AC ABZ11345;  
XX  
DT 20-JAN-2003 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 227.  
XX  
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cytostatic; immunomodulator; neutropenic; dermatological;  
KW antiparkinsonian; antididiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
KW antiarthritic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200270539-A2.  
XX  
PD 12-SEP-2002.  
XX  
PF 05-MAR-2002; 2002WO-US05095.  
XX  
PR 05-MAR-2001; 2001US-0799451.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX  
DR WPI: 2002-759812/82.  
DR P-PSDB; ABP69127.  
XX  
PT New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
PT platelet or coagulation disorders  
XX  
PS Claim 1; SEQ ID NO 227; 1012pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences  
CC (ABZ11119-ABZ12066) or their mature protein coding portion, active domain  
CC nucleotide sequence selected from any of 948 sequences  
CC (ABZ11119-ABZ12066) or their mature protein coding portion, active domain



Matches 437; Conservative 0; Mismatches 346; Indels 33; Gaps 4;

QY 222 CCGCTTGTCTCGCGACAGCGAGCCCTTCTGCCACCGCACTCCAGCGCGCGCGCAGCC 281  
 Db |||||  
 QY 225 CCGACCCAGTGTGCGCGCGACTTCCCGTGGAGCTGGACCCGACTGGTCCCCCAGGCC 284  
 Db |||||  
 QY 282 CGAGCCCGCGCGCGCTCGCCAGCTATCCGAGCTGCGCCCTCTTTCGGGGGCGCTGCTCG 341  
 Db |||||  
 QY 285 GGCCAGGCTCGCGCGCGCGCGCTCGCGACCTGAGCTTCTTTCGGGGGCGCTTCTCG 344  
 Db |||||  
 QY 342 CCGCGCGCAGCTCTCAAGCGCTCGAGCAGCGCGCTGCGAGCTTCCGCGAGTCCCGACC 401  
 Db |||||  
 QY 345 TCGCGTCTGCTGCGCGCGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 395  
 Db |||||  
 QY 402 CAGCGCGAGTGTCTCGCGAGCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 461  
 Db |||||  
 QY 396 CAGCGAAGAGATGGAGTGGAGTTCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 455  
 Db |||||  
 QY 462 TTACTTCAAGGCAATATCTCCCAAGCCATCGCGCTGCTCACACTTCTTCTACTGAA 521  
 Db |||||  
 QY 456 CTACTTCAGATCAACAAGTGGAGAGCTGTGTCGACGACACACCTTCTTCTCGTGG 515  
 Db |||||  
 QY 522 GCATCTGTATGACGAATGATGAAGAGAACATGCGCATATTATAAGAGCTGCTGTGTG- 580  
 Db |||||  
 QY 516 CAATCTCTGACCATGGAATGACGAGAACCTAGACTATTACCAAAACCATGCTGGAGT 575  
 Db |||||  
 QY 581 --CGAGGACTATTAAGACCTGGAACCAAGTCATATGAAGCCCTGTTTATCCGAGC 638  
 Db |||||  
 QY 576 GAAGGAGCGGCTTCAAGATCTTGAGACTCAACCCCATATGCAAGAAATTCGACTGG 635  
 Db |||||  
 QY 639 AGTGGCGGATCAACCGGTGAGAACTGGAGAACATCCATCAGACATGGAGCTGGCCCT 698  
 Db |||||  
 QY 636 AGTGGCACTTACTCAGAGGAACAGCCACAGGAAGCTGTGCCACCTAGAGCGCGCT 695  
 Db |||||  
 QY 699 TCCGACTTCTCAAGCCTTTACAGTGTCTCGAGCTGCGAGGTTCCAGGAGAT 758  
 Db |||||  
 QY 696 GCAAGAAATCTTGTGCGCTATGAGGAGTCCCTGCGGAGGCGCTTATGACTA 755  
 Db |||||  
 QY 759 CAAGGACTTCAA-----GGATTCTTACCTTTCCATAGCAGATCAATTA 800  
 Db |||||  
 QY 756 CGATGGCTACACTACCTTGATGACACGCTGACCTTCTCCAGGCGCATCAGAGATCA 815  
 Db |||||  
 QY 801 TGTAGAGTCTGGAATGCAATATCAGTGTGAGGAGACTT--CACCCAGTTATAGG 857  
 Db |||||  
 QY 816 CATCCAGTCTCACTGTGAAGCAGAACTGTGTCAGGAGCTGTCTCCACCCAGTCG 875  
 Db |||||  
 QY 858 AGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTTACTTGCATTTGCTTATTA 917  
 Db |||||  
 QY 876 AGAAGACCTTTGAAGACTTCTCCCATCGCATTAATATATCTGCAAGTTTGCCTACTA 935  
 Db |||||  
 QY 918 TAAGTTGAACGCTGAGATGACAGCCCGCTGTGCACTCAGCTATCTGCTTTGATCA 977  
 Db |||||  
 QY 936 TAACATTTGGAAATATACAGGCTGTGTAATGTGCAAGACTATCTTCTTCTTCTCC 995  
 Db |||||  
 QY 978 GAATGACAGGTTCATGACAGCAACCTTGTGTATTA 1013  
 Db |||||  
 QY 996 CAATGACAGGTGTATGAACCAAAATTTGGCTATTA 1031  
 Db |||||

## RESULT 12

ABZ11346

ID ABZ11346 standard; cDNA; 2242 BP.

XX

AC

XX

XX

DT

XX

DE

XX

KW

KW

KW

KW

KW

KW

arthritis; cytostatic; immunomodulator; neotropic; neuroprotective;  
 anti-parkinsonian; antididiabetic; immunosuppressive; dermatological;  
 haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoicide;  
 antiarthritic; gene; ss.  
 OS Homo sapiens.  
 PN WO200270539-A2.  
 PD 12-SEP-2002.  
 XX 05-MAR-2002; 2002WO-US05095.  
 XX 05-MAR-2001; 2001US-0799451.  
 PR (HYSE-) HYSEQ INC.  
 XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX WPI; 2002-759812/82.  
 DR P-PSDB; ABP69129.  
 XX New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative, or  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
 PT platelet or coagulation disorders  
 XX Claim 1; SEQ ID NO 228; 1012pp + Sequence Listing; English.  
 PS The invention relates to an isolated polynucleotide (I) comprising a  
 CC nucleotide sequence selected from any of 948 sequences  
 CC (ABZ11119-ABZ12066) or their mature protein coding portion, active domain  
 CC coding protein or complementary sequences. The polynucleotides are useful  
 CC for identifying expressed genes or for physical mapping of human genome.  
 CC The encoded polypeptides (ABP6902-ABP6949) are useful as molecular  
 CC weight markers, as a food supplement, for generating antibodies, in  
 CC medical imaging, screening and diagnostic assays and for treating  
 CC cell-proliferative disorders (cancer), neurodegenerative diseases  
 CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
 CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
 CC disorders, platelet or coagulation disorders, wound, burns, incision,  
 CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
 CC parasitic), arthritis, etc.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 2242 BP; 518 A; 655 C; 625 G; 444 T; 0 other;

Query Match 13.0%; Score 156.4; DB 24; Length 2242;

Best Local Similarity 53.6%; Pred. No. 1.2e-24;

Matches 437; Conservative 0; Mismatches 346; Indels 33; Gaps 4;

QY 222 CCGCTTGTCTCGCGACAGCGAGCCCTTCTGCCACCGCACTCCAGCGCGCGCGCAGCC 281

Db 256 CCGCACCCAGTGTGCGCGCGACTTCCCGTGGAGCTGGACCCCGCTGCTCCCCCAGCCC 315

QY 282 CGAGCCCGCGCGCGCTTCCGAGCTATCCGAGCTGCGCTCTTTCGGGGCGCTGCTGCG 341

Db 316 GGGCCAGGCTCGGGCGCGCGCGCTGCGGAGCTGAGCTTCTTTCGGGGCGCTTCTGCG 375

QY 342 CCGCGCGCACTGCTCAAGCGGTGCAAGCAGGGCTTCCAGCCTTCCGCCAGTCCCGAGCC 401

Db 376 TCGCGTGTGCTGCTGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426

QY 402 CAGCCCGCGAGTGTCTGGCGGACTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 461

Db 427 CAGCGAAGAGATGGAGCTGGAGTTCGCCAAGCGGAGCGCGCGCGCGCGCGCGCGCG 486

QY 462 TTACTTCAAGGCAATATATCTCCCAAGCCATCGCGCTGCTCACACCTTTCTTACTGAA 521

Db

QY

Human polynucleotide SEQ ID NO 228.

Human; genome mapping; gene therapy; food supplement; virus; fungus;

cell-proliferative disorder; neurodegenerative disease; bacterial;

Parkinson's disease; Alzheimer's disease; autoimmune disease;

multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;

487 CTACTTCAAGATCAACAACTTGGAGAAAGCTGTGCTGCAGACACACACCTTCTTCGTGGG 546  
QY 522 GCATCTGTGATGCGAAATGATGAAGGACATGCGCATATTATAGAGCTTCTGCTGTG- 580  
Db 547 CAATCTGTGACATCGGAATGCGAGACCTAGACTATTACCAACCATGCTGTGAGT 606  
QY 581 --CCGAGGACTACATTAAAGACCTGGAACCAAGTCAATATGAAGCCCTGTTTCATCCGAGC 638  
Db 607 GAAGGAGGCGGACTTCAAGGATCTTGAAGTCAACCCCATATGCAAGAAATTTGACTGGG 666  
QY 639 AGTGGGGCATACCAAGGTGAGAACTGGAGACATCCATCAGACATGAGCTGGCCCT 698  
Db 667 AGTGGACTTCTACTCAGAGAACCCACAGAGAGCTGTGCCACCTAGAGGGGGGCT 726  
QY 699 TCCGACTTCTTCAAGGCTTTTACAGTGTCTGCGAGCTGCGAGGGTTCAGAGGAGAT 758  
Db 727 GCAAGAAATACTTTGTGGCTATGAGGAGTGGCGTCCCTCTGCGAAGGGCCCTATGACTA 786  
QY 759 CAGGACTTCAA-----GGATTCTACCTTCCATAGCAGATCATTA 800  
Db 787 CAGTGGCTTAACTACTTGTAGTACAGCTGACCTTCCAGGCCATCAGATCATTA 846  
QY 801 TGTAGAGTCTGGAATCAAAATACAGTGTGAAGAGAACCT---CACCCGAGTTATAGG 857  
Db 847 CATCAGGTCCTCAACTGTAGCAGACAGTGTGTCAAGGAGCTTCTTCCACCCAGTCG 905  
QY 858 AGGCTATCGGTGAGAAATTTGTGGTACCATGTATCATTTACTTGCAGTTGCTTATTA 917  
Db 907 AGAGAGCCCTTTGAAGACTTCTCCCATCGCATTAATATCTGCAGTTGCTTACTA 966  
QY 918 TAAGTTGAACGACCTGAAGAATGACGCCCCCTGTGCACTGATCTGTCTTTGATCA 977  
Db 967 TAACATGGGAATTTATACAGAGCTGTGTAATGTGCGAAGACCTATCTTCTTCTTCCC 1026  
QY 978 GAATCAGAGGTCATGACGAGACCTGCTTATTA 1013  
Db 1027 CAATCAGAGGTCATGAACCAAAATTTGGCCTATTA 1062

RESULT 13  
AAH14513  
ID AAH14513 standard; cDNA; 2524 BP.  
XX  
AC AAH14513;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:12044.  
XX  
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
DE Homo sapiens.  
XX  
FN BP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116125.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602

full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -  
Claim 8; SEQ ID 12044; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to a complementary strand of a polynucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the polynucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH19742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2524 BP; 583 A; 747 C; 704 G; 490 T; 0 other;

Query Match  
Best Local Similarity 13.08; Score 156.4; DB 22; Length 2524;  
Matches 437; Conservative 0; Mismatches 346; Indels 33; Gaps 4;  
QY 222 CCGCTTGTGCGCAGCAGGAGGCTTCTGCGACCGCACTGCGCGCGCGCGCGCGC 281  
Db 240 CCGCACCAGTGTGCGCGCGACTTCCGCTGGAGCTGGACCCCGGACTGCTCCCGCAGCC 299  
QY 282 CGAGCCGCGCGCGCGCTCCCGAGCTATCCGAGCTGCGCTCTTCGGGGGCTGCTGG 341  
Db 300 GCGCAGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 359  
QY 342 CCGCGCGCCTGCTCAAGCGCTGCAAGAGGCGCTTCCGAGCTTCCGCGAGTCCCGCAGC 401  
Db 360 TCGCGCTGCTGCTGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 410  
QY 402 CAGCGCGAGGTGCTGCGGAGCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 461  
Db 411 CAGCGAGAGATGAGCTGAGGTTCGCAAGCGGAGCGCGCGCGCGCGCGCGCGCG 470  
QY 462 TTACTTCAAGGCAAAATATCTCCCGAAAGCCATTCGCGGTGTGTACACCTTTCTACTGAA 521  
Db 471 CTACTTCAAGATCAACAAAGTTGGAGAAAGCTGTGCTGCGAGCAGACACACCTTTCTG 530  
QY 522 GCATCTGTATGCGAAATGATGAAGGAGACATGCGCATATTATAGAGCTGCGCTGGTG- 580  
Db 531 CAATCTGAGCAGATGGAATGAGAGAACCTAGACTATTACCAACCATGCTGTGAGT 590  
QY 581 --CCGAGGACTACATTAAAGACCTGGAACCAAGTCAATATGAAGCCCTGTTTCATCCGAGC 638  
Db 591 GAAGGAGGCGGACTTCAAGGATCTTGAAGACTCAACCCCATATGCAAGAAATTTGACTGGG 650  
QY 639 AGTGGGGCATACAAAGGTGAGAACTGGAGAACATCCATCAGACATGAGCTGGCCCT 698  
Db 651 AGTGGGACTTACTCAGAGGAACAGCCACAGGAGGTGTGCCCGCCCTAGAGGCGGCT 710  
QY 699 TCCGAGCTTCTTCAAGCCCTTTTACAGTGTCTGCGAGCTGCTGCGAGGTTCCAGGAGAT 758  
Db 711 GCAAGAAATACTTTGTGGCCTATGAGGAGTGGCGTCCCTCTGCGAAGGGCCCTATGACTA 770  
QY 759 CAAGGACTTCAA-----GGATTCTTACCTTTCCATAGCAGATCATTA 800







AAH14589 standard; cDNA; 2585 BP.  
AAH14589;  
26-JUN-2001 (first entry)  
Human cDNA sequence SEQ ID NO:12193.  
Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
Homo sapiens.  
EP1074617-A2.  
07-FEB-2001.  
28-JUL-2000; 2000EP-0116126.  
29-JUL-1999; 93JP-0248036.  
27-AUG-1999; 95JP-0300253.  
11-JAN-2000; 2000JP-0118776.  
02-MAY-2000; 2000JP-0183767.  
09-JUN-2000; 2000JP-0241899.  
(HELI-) HELIX RES INST.  
Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
WPI; 2001-318749/34.  
Primer sets for synthesizing polynucleotides, particularly the 5602  
full-length cDNAs defined in the specification, and for the detection  
and/or diagnosis of the abnormality of the proteins encoded by the  
full-length cDNAs -  
Claim 8; SEQ ID 12193; 2537pp + CD ROM; English.  
The present invention describes primer sets for synthesizing 5602  
full-length cDNAs defined in the specification. Where a primer set  
comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
to the complementary strand of a polynucleotide which comprises one of  
the 5602 nucleotide sequences defined in the specification, where the  
oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
of an oligonucleotide comprising a sequence complementary to the  
complementary strand of a polynucleotide which comprises a 5'-end  
sequence and an oligonucleotide comprising a sequence complementary to a  
polynucleotide which comprises a 3'-end sequence, where the  
oligonucleotide comprises at least 15 nucleotides and the combination of  
the 5'-end sequence/3'-end sequence is selected from those defined in  
the specification. The primer sets can be used in antisense therapy and  
in gene therapy. The primers are useful for synthesizing polynucleotides,  
particularly full-length cDNAs. The primers are also useful for the  
detection and/or diagnosis of the abnormality of the proteins encoded by  
the full-length cDNAs. The primers allow obtaining of the full-length  
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
represent oligonucleotides, all of which are used in the exemplification  
of the present invention.  
Sequence 2585 BP; 590 A; 759 C; 723 G; 513 T; 0 other;

Search completed: January 30, 2004, 08:32:16  
Job time : 317.764 secs

Query Match 13.0%; Score 156.4; DB 22; Length 2585;  
Best Local Similarity 53.6%; Pred. No. 1.2e-24;  
Matches 437; Conservative 0; Mismatches 346; Indels 33; Gaps 4;  
222 CCGCTTCTGCGGACAGCGAGGCTTCTCCACCGCAACTGACGAGCGCGCGGAGCG 281  
277 CCGCACCCAGTGTGCGCGCCGCACTTCCGCTGGAGTGGACCCCGACTGTGTCGCCAGCCC 336  
282 CGAGCCGCGCGCGGCTCGCGAGCTATCCGAGCTGCGCTCTTGGGGGCTGTCTGG 341

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 05:50:18 ; Search time 1006.41 Seconds  
(without alignments)  
10382.969 Million cell updates/sec

Title: US-09-729-674-1  
Perfect score: 3871  
Sequence: 1 ttctctctccctccctttt.....ataaaaaaaaaaaaaaaaaa 3871

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2004.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3870.2	100.0	3871	20	Human secreted pro
2	3870.2	100.0	3871	22	Human cDNA encodin
3	3870.2	100.0	3871	24	Human polynucleoti
4	2051.2	53.0	2549	21	Human prostate can
5	1955.4	50.5	1951	22	Human cDNA sequenc
6	1474.4	38.1	1586	21	Human colon cancer
7	1062	27.4	2499	23	Human prostate exp
8	1062	27.4	2499	23	Human prostate exp

9	695.6	18.0	818	22	Human cDNA clone (
10	587.4	15.2	660	24	Human colon cancer
11	466.4	12.0	618	21	Human colon cancer
12	465.8	12.0	2347	24	Oesophagus cancer
13	406.4	10.5	408	21	Human colon cancer
14	392.4	10.1	394	24	Human colon cancer
15	371.6	9.6	401	24	Gene #1453 used to
16	369.4	9.5	400	16	Human tumour assoc
17	355.6	9.2	407	22	Human gene signatu
18	340	8.8	351	22	Human tumour assoc
19	337.2	8.7	373	21	Human cDNA clone (
20	331.2	8.6	595	23	Human secreted exp
21	267.6	6.9	350	23	Human prostate exp
22	263.4	6.8	274	24	Human prostate exp
23	258.6	6.7	279	23	Lung cancer relate
24	241.2	6.2	43545	24	cDNA #683 encoding
25	239.2	6.2	6670	22	Human cadherin-like
26	239.2	6.2	7713	22	Human reproductive
27	238.4	6.2	8658	22	Human reproductive
28	238	6.1	295	22	Human immune/haema
29	235.4	6.1	32351	21	Human immune/haema
30	235.4	6.1	32351	21	Human adenosine re
31	235.4	6.1	40298	21	Human low adenosin
32	235.4	6.1	40298	21	Human adenosine re
33	234.6	6.1	4275	22	Human nervous syst
34	234.6	6.1	4883	22	Human immune/haema
35	234.6	6.1	8894	23	Human immune/haema
36	234.6	6.1	16997	22	Genomic sequence #
37	234.6	6.1	16997	22	Human immune/haema
38	234.2	6.1	9163	24	Human immune/haema
39	234.2	6.1	97835	24	Human endothelial
40	233.8	6.0	9620	22	Human cDNA differe
41	233.6	6.0	43950	24	Human kinase genom
42	233.4	6.0	6138	22	Human liver cell s
43	233.2	6.0	763	22	Human cDNA clone (
44	232.6	6.0	2263	22	Human cDNA sequenc
45	232.6	6.0	5979	22	Human immune/haema

## ALIGNMENTS

## RESULT 1

AAK60801

ID AAK60801 standard; DNA; 3871 BP.

XX AAK60801;

AC XX 09-AUG-1999 (first entry)

DT XX Human secreted protein encoding DNA (clone bd306-7).

DE XX Secreted protein; kidney; lung; brain; blood; testis; bone marrow;

KW nutritional activity; cytokine; cell proliferation; immune stimulation;

KW hematopoiesis regulation; tissue growth; thrombolytic; gene therapy;

KW anti-inflammatory; tumour invasion; ss.

OS Homo sapiens.

PN WO9926961-A1.

XX -03-JUN-1999.

PD XX 24-NOV-1998; 98WO-US25149.

PF XX 23-NOV-1998; 98US-0197886.

PR XX 26-NOV-1997; 97US-0066804.

XX (GENY) GENETICS INST INC.

FA Agostino MJ, Clark HF, Collins-Racie LA, Evans C;

PI Fechtel K, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PI Steininger RJ, Treacy M, Wong GG;



1681 DB ||||| CCAGGCTGCTCGAACTCTTGACCTCAGATGATCCATCTGCTTGGCCCTCCACAGTGC 1740  
1741 QY ||||| TGGGATTTACAGCGTGGAGCCACCATGCGCGGCTCTTTCTCACCTTTACACCTGCTCTCT  
1741 DB ||||| TGGGATTTACAGCGTGGAGCCACCATGCGCGGCTCTTTCTCACCTTTACACCTGCTCTCT  
1801 QY ||||| TATCTCAGATCTGTTTTCACACCTTCATPCCCTGCTCTTCTCCTCAGTGTGACACTGCTTC  
1801 DB ||||| TATCTCAGATCTGTTTTCACACCTTCATPCCCTGCTCTTCTCCTCAGTGTGACACTGCTTC  
1861 QY ||||| CCCATGTTTCCAGAAAATCAGTATATATTTTTTAAATAGAAAATCCTCTAGAA 1980  
1861 DB ||||| TTTTTTGTGTTTCCAGAAAATCAGTATATATTTTTTAAATAGAAAATCCTCTAGAA 1980  
1981 QY ||||| GATGATAATTTGAAAACCTCTTGGCTTATTTGCTTTTCCAGATTTTGTCTCTCTTC 2040  
1981 DB ||||| GATGATAATTTGAAAACCTCTTGGCTTATTTGCTTTTCCAGATTTTGTCTCTCTTC 2040  
2041 QY ||||| TCCCATCCGGAAGATGTTGGAAGACATAGGCTAAATTTCTCCAGCTCACAAATGCTC 2100  
2041 DB ||||| TCCCATCCGGAAGATGTTGGAAGACATAGGCTAAATTTCTCCAGCTCACAAATGCTC 2100  
2101 QY ||||| TTTCACTTGTCTGACTTGTACCAATTTAGCAACCTCTGCTTATTTGCTTTTCCAGATTTTGTCTCTCTTC 2160  
2101 DB ||||| TTTCACTTGTCTGACTTGTACCAATTTAGCAACCTCTGCTTATTTGCTTTTCCAGATTTTGTCTCTCTTC 2160  
2161 QY ||||| TGTAGAGTGCAGAAATGTTGGCTTTTGGCCCTTGTGCTTCTCCAAATTTCAACGGTGG 2220  
2161 DB ||||| TGTAGAGTGCAGAAATGTTGGCTTTTGGCCCTTGTGCTTCTCCAAATTTCAACGGTGG 2220  
2221 QY ||||| CCGATCCCATTTGAGGACAACTGTTAGTTATAGCTCCGAGTCTGGAAGGAGAAAGC 2280  
2221 DB ||||| CCGATCCCATTTGAGGACAACTGTTAGTTATAGCTCTCGAGTCTGGAAGGAGAAAGC 2280  
2281 QY ||||| CAGAGCTGTCTAGTTTTCATTCTCTTTTCACTTAAATATTTATGAGTACCTTCTGCTGC 2340  
2281 DB ||||| CAGAGCTGTCTAGTTTTCATTCTCTTTTCACTTAAATATTTATGAGTACCTTCTGCTGC 2340  
2341 QY ||||| TAGGCATTCAGCTTGGGAACTAGAGTACTTACAGAAATACAGGGAAGTTCCTGCTGCT 2400  
2341 DB ||||| TAGGCATTCAGCTTGGGAACTAGAGTACTTACAGAAATACAGGGAAGTTCCTGCTGCT 2400  
2401 QY ||||| CATGGAGCTTACATTTACAGGAGAAAGAGATAGCCAAATACATAGGAATAATATATATAC 2460  
2401 DB ||||| CATGGAGCTTACATTTACAGGAGAAAGAGATAGCCAAATACATAGGAATAATATATATAC 2460  
2461 QY ||||| AAGGTATCATGTAGTGAATTTGCTGTGAGAAATTAAGCAGGAGGAGGAGTAAGAA 2520  
2461 DB ||||| AAGGTATCATGTAGTGAATTTGCTGTGAGAAATTAAGCAGGAGGAGGAGTAAGAA 2520  
2521 QY ||||| TCTGAGATGAGGCTGCACTTTTAAATGCGGCTTCACTGGGAAATGTGACGTTGAGCAGA 2580  
2521 DB ||||| TCTGAGATGAGGCTGCACTTTTAAATGCGGCTTCACTGGGAAATGTGACGTTGAGCAGA 2580  
2581 QY ||||| GAGGTTAGGAGTGTGATCTCGACAGGCTTCCAGGACAGGAGAACAGATGTGCACTG 2640  
2581 DB ||||| GAGGTTAGGAGTGTGATCTCGACAGGCTTCCAGGACAGGAGAACAGATGTGCACTG 2640  
2641 QY ||||| CCCAAAGTGTGAGTGTCTCTACGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGTGC 2700  
2641 DB ||||| CCCAAAGTGTGAGTGTCTCTACGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGTGC 2700  
2701 QY ||||| TGGGACAGGAGTGAATGGAAGAGAGGCGGCTTGGGAGGAGCAGGTTGTTGAGGAGGAGGAG 2760  
2701 DB ||||| TGGGACAGGAGTGAATGGAAGAGAGGCGGCTTGGGAGGAGCAGGTTGTTGAGGAGGAGGAG 2760  
2761 QY ||||| CTTCTGCTAAGTGAATGGAACCACTGGAGGTTTGAACAGAGAGTGCCTTGAATGAT 2820

2761 DB ||||| CTTCTGCTAAGTGAATGGAACCACTGGAGGTTTGAACAGAGAGTGCCTTGAATGAT 2820  
2821 QY ||||| TTATATTTTGAAGGCTCATTTCTAGCTGCAATATTTGTGAATAAATTTAGTGGACAGGGC 2880  
2821 DB ||||| TTATATTTTGAAGGCTCATTTCTAGCTGCAATATTTGTGAATAAATTTAGTGGACAGGGC 2880  
2881 QY ||||| AGAAGGAAGAGGAGAGACCTGTTAGGAAGCTACTGCAAGGTTTCCAGGCTTGGGCTGGGC 2940  
2881 DB ||||| AGAAGGAAGAGGAGAGACCTGTTAGGAAGCTACTGCAAGGTTTCCAGGCTTGGGCTGGGC 2940  
2941 QY ||||| CACACCAACAGCAGTGTGCAATATCTAGATTTATTTTGAATAAAGAGCAATAGGATTTGC 3000  
2941 DB ||||| CACACCAACAGCAGTGTGCAATATCTAGATTTATTTTGAATAAAGAGCAATAGGATTTGC 3000  
3001 QY ||||| TGAGATTTGAATGTTGAGTGTAAAGAGAGAGAGATTAATGATGACATTAAGGTTTTC 3060  
3001 DB ||||| TGAGATTTGAATGTTGAGTGTAAAGAGAGAGAGATTAATGATGACATTAAGGTTTTC 3060  
3061 QY ||||| GCTCAATAGCAGGAGAGATGAGTGTACCACTTACTGAATAGGAAGGATGGGCTGGGT 3120  
3061 DB ||||| GCTCAATAGCAGGAGAGATGAGTGTACCACTTACTGAATAGGAAGGATGGGCTGGGT 3120  
3121 QY ||||| AAGTAAAGAAATTTGTCGAAAGCAGGCTGCTGTGTTGGAATGGAAGGATTTGCTGCTGCA 3180  
3121 DB ||||| AAGTAAAGAAATTTGTCGAAAGCAGGCTGCTGTGTTGGAATGGAAGGATTTGCTGCTGCA 3180  
3181 QY ||||| AATCAAACTGAGATTTCTCTCAGGCTCAGGCTCTGAGCAGAGCTCGAGACAGGATCTGAA 3240  
3181 DB ||||| AATCAAACTGAGATTTCTCTCAGGCTCAGGCTCTGAGCAGAGCTCGAGACAGGATCTGAA 3240  
3241 QY ||||| TGCATCTGCTTTATTTGTTGGGGTCTCTCAGAAAGGAACTGTGAAAGCCTTTATCAGTC 3300  
3241 DB ||||| TGCATCTGCTTTATTTGTTGGGGTCTCTCAGAAAGGAACTGTGAAAGCCTTTATCAGTC 3300  
3301 QY ||||| ATTTATTTGCTGTGAGAAATTTCTCTGAGGCTGCTGAGTACATTTGAAAGCAGGATCTTCA 3360  
3301 DB ||||| ATTTATTTGCTGTGAGAAATTTCTCTGAGGCTGCTGAGTACATTTGAAAGCAGGATCTTCA 3360  
3361 QY ||||| GTTGAAGGCAAGTCTCTGGAAGAGAGGCTGTAGGCACTCTGAGCAGTACCTTGGGTGAG 3420  
3361 DB ||||| GTTGAAGGCAAGTCTCTGGAAGAGAGGCTGTAGGCACTCTGAGCAGTACCTTGGGTGAG 3420  
3421 QY ||||| TGTGTTGGGGTGGGGTCTCTGAGGCACTGCTGTGTAAGGATCTGAGGAGGACCCACA 3480  
3421 DB ||||| TGTGTTGGGGTGGGGTCTCTGAGGCACTGCTGTGTAAGGATCTGAGGAGGACCCACA 3480  
3481 QY ||||| GCGCCCTTACTGAAACCACTCAGCATGTGAGTGTGCAATTTAAAGCCATGAGCTGAGGGC 3540  
3481 DB ||||| GCGCCCTTACTGAAACCACTCAGCATGTGAGTGTGCAATTTAAAGCCATGAGCTGAGGGC 3540  
3541 QY ||||| CACTGAGATTTGCTCTGAGTATTTACTGAGAGCAACAGAAAGAGCCATGATGAGGCC 3600  
3541 DB ||||| CACTGAGATTTGCTCTGAGTATTTACTGAGAGCAACAGAAAGAGCCATGATGAGGCC 3600  
3601 QY ||||| TTGGGCTCTCTGGGAAATGGAATCAGCCAAAGACTGAGAGGATTTACCTTAAGGTC 3660  
3601 DB ||||| TTGGGCTCTCTGGGAAATGGAATCAGCCAAAGACTGAGAGGATTTACCTTAAGGTC 3660  
3661 QY ||||| AGAGAAAACCAAGAGAGTGTGTTTCTGGAAGCTGAGCTTTCTTATTTCAACCTCATTC 3720  
3661 DB ||||| AGAGAAAACCAAGAGAGTGTGTTTCTGGAAGCTGAGCTTTCTTATTTCAACCTCATTC 3720  
3721 QY ||||| CTTCTCAGAAATAGGCTTGTGTTGTTGGGCTTCCAGGCTTGAAGGAGAGAGAGA 3780  
3721 DB ||||| CTTCTCAGAAATAGGCTTGTGTTGTTGGGCTTCCAGGCTTGAAGGAGAGAGAGA 3780  
3781 QY ||||| AAGGACAGGCTTTGGGAAAACAGACTTTTCTGCAATAGCTGGGAGGAAATAAGGA 3840  
3781 DB ||||| AAGGACAGGCTTTGGGAAAACAGACTTTTCTGCAATAGCTGGGAGGAAATAAGGA 3840  
3841 QY ||||| TAGAGTGTATAAATAAAAAAAAAAAAAA 3871  
3841 DB ||||| TAGAGTGTATAAATAAAAAAAAAAAAAA 3871



QY 961 ACTTGAGTGGCTGCTATTATATAGTTGAACGACCTGAAGAATGACGCCCTCTGTGAGTCA 1020  
DB 961 ACTTGAGTGGCTGCTATTATAGTTGAACGACCTGAAGAATGACGCCCTCTGTGAGTCA 1020  
QY 1021 GCTATCTGCTCTTTGATCAGAAATGACAAAGGTGATGACGACGAACTGGTGTATTACCACT 1080  
DB 1021 GCTATCTGCTCTTTGATCAGAAATGACAAAGGTGATGACGACGAACTGGTGTATTACCACT 1080  
QY 1081 ACCACAGGAGACATCTGGGCGCTCTCGGATGAGCACTTCCAGCCAGACCTGAACAGTTC 1140  
DB 1081 ACCACAGGAGACATCTGGGCGCTCTCGGATGAGCACTTCCAGCCAGACCTGAACAGTTC 1140  
QY 1141 AGTCTTTAAATGTGACCACTCCAGAAAGGAGTGTATGATCTTGAAGAAATATAA 1200  
DB 1141 AGTCTTTAAATGTGACCACTCCAGAAAGGAGTGTATGATCTTGAAGAAATATAA 1200  
QY 1201 TGGATGATGAGGAGAAAGTTGGGAATATGAGATGACCTCTTGGAACTGGAGGAGA 1260  
DB 1201 TGGATGATGAGGAGAAAGTTGGGAATATGAGATGACCTCTTGGAACTGGAGGAGA 1260  
QY 1261 CCAGTACGCCACACAGCAACCAAGAGACTTCTCTTGGCGTTACAGGAAACACAGATTCTT 1320  
DB 1261 CCAGTACGCCACACAGCAACCAAGAGACTTCTCTTGGCGTTACAGGAAACACAGATTCTT 1320  
QY 1321 TGTCTTTTCCAAACAGCCAGGCTGTGATPACCTCAGAGCCTTCTCTTTACTCTCCAAA 1380  
DB 1321 TGTCTTTTCCAAACAGCCAGGCTGTGATPACCTCAGAGCCTTCTCTTTACTCTCCAAA 1380  
QY 1381 GTGAAAGGAGAGCCCGTCTCTTAAGTGCATGTATCAGGGGTGAGCCTGCTTTCTT 1440  
DB 1381 GTGAAAGGAGAGCCCGTCTCTTAAGTGCATGTATCAGGGGTGAGCCTGCTTTCTT 1440  
QY 1441 ATCTTACACCTGCCACTCATGTTTACACCTATCTTTCTCACCTTTTTCAGATGGA 1500  
DB 1441 ATCTTACACCTGCCACTCATGTTTACACCTATCTTTCTCACCTTTTTCAGATGGA 1500  
QY 1501 GTCTCGTCTCTTCCAGCTGAGTGCAATGCGCATGTTCTCAGCTCACTGCAACCTCC 1560  
DB 1501 GTCTCGTCTCTTCCAGCTGAGTGCAATGCGCATGTTCTCAGCTCACTGCAACCTCC 1560  
QY 1561 GCCTCTTGGGTTCAAGCAATCTGCTCATCAGCTCCGAGTACCTGGGATACAGGCA 1620  
DB 1561 GCCTCTTGGGTTCAAGCAATCTGCTCATCAGCTCCGAGTACCTGGGATACAGGCA 1620  
QY 1621 TGTGCCACACGCGCGCTAAATTTGTAATTTTAGTAGACGCGGTTTGCCATGTTGG 1680  
DB 1621 TGTGCCACACGCGCGCTAAATTTGTAATTTTAGTAGACGCGGTTTGCCATGTTGG 1680  
QY 1681 CCAGGCTGGTCTGAACTCTGACTCAGATGATCCATCTGCGCTTGGCCTCCACAGTGC 1740  
DB 1681 CCAGGCTGGTCTGAACTCTGACTCAGATGATCCATCTGCGCTTGGCCTCCACAGTGC 1740  
QY 1741 TGGGATACAGGCTGAGCCACCATGCCGCGCTCTTTCTCACCTTTACACCTGTCTCT 1800  
DB 1741 TGGGATACAGGCTGAGCCACCATGCCGCGCTCTTTCTCACCTTTACACCTGTCTCT 1800  
QY 1801 TATCCTCAGATCTGTTTTCACACTTTCACCTTCACTCCCTGCTCTCTCATGTTTCATGTTCTC 1860  
DB 1801 TATCCTCAGATCTGTTTTCACACTTTCACCTTCACTCCCTGCTCTCTCATGTTTCATGTTCTC 1860  
QY 1861 CCCATGTTTATAGCTGCTCTTCTTACATTTGGTTTGAAGGCGACTTCTCTCTGCTTG 1920  
DB 1861 CCCATGTTTATAGCTGCTCTTCTTACATTTGGTTTGAAGGCGACTTCTCTCTGCTTG 1920  
QY 1921 TTTTCTGTTTTCAGAGAAATCAGTATATTTTAAATAGAAAAACATTCCTCTAGAA 1980  
DB 1921 TTTTCTGTTTTCAGAGAAATCAGTATATTTTAAATAGAAAAACATTCCTCTAGAA 1980  
QY 1981 GATGATAATGTGAAACCTCTCTTTCGCTTATTTGCTTTTCCAGATTTTAGTCTCTCTT 2040  
DB 1981 GATGATAATGTGAAACCTCTCTTTCGCTTATTTGCTTTTCCAGATTTTAGTCTCTCTT 2040

QY 2041 TCCCATCCGGGAAAGATGGTGAAGACATAGGCTAAATTTCTCCAGCCTCAATGGTC 2100  
DB 2041 TCCCATCCGGGAAAGATGGTGAAGACATAGGCTAAATTTCTCCAGCCTCAATGGTC 2100  
QY 2101 TTCACTTGGTCTTGACTTTTACCAATTTCTAGCACCACTGAAAAACAAGTTGAGTAGAG 2160  
DB 2101 TTCACTTGGTCTTGACTTTTACCAATTTCTAGCACCACTGAAAAACAAGTTGAGTAGAG 2160  
QY 2161 TGTAGAGTGCAGAAATGTGGCTTTTGGCCCACTTTGCACTCTCCAAATTTACAAAGCTTGG 2220  
DB 2161 TGTAGAGTGCAGAAATGTGGCTTTTGGCCCACTTTGCACTCTCCAAATTTACAAAGCTTGG 2220  
QY 2221 CCGATCCCAATTTGAGGACAACTGATTTAAGTCTCCGAGTTGGAAAGGAAAGC 2280  
DB 2221 CCGATCCCAATTTGAGGACAACTGATTTAAGTCTCCGAGTTGGAAAGGAAAGC 2280  
QY 2281 CAGAGCTGTAGTTTTCATTCTTTCAGTAAATATTTATTGAGTACCTACTCTGTGC 2340  
DB 2281 CAGAGCTGTAGTTTTCATTCTTTCAGTAAATATTTATTGAGTACCTACTCTGTGC 2340  
QY 2341 TAGGCATTTGACCTGGGAACTAGAGATCTTACAGAAATACAGGGGAAAGTTCCCTGTGC 2400  
DB 2341 TAGGCATTTGACCTGGGAACTAGAGATCTTACAGAAATACAGGGGAAAGTTCCCTGTGC 2400  
QY 2401 CATGGAGCTTACATTTCTACAGGAGAAAGATAGCCCAATACATAGGAATTAATATATAC 2460  
DB 2401 CATGGAGCTTACATTTCTACAGGAGAAAGATAGCCCAATACATAGGAATTAATATATAC 2460  
QY 2461 AAGGTATCATGTPAGTGAATTTGCTGTGAGAAAAATTAAGCAGGGGAGGAGTAAGAAA 2520  
DB 2461 AAGGTATCATGTPAGTGAATTTGCTGTGAGAAAAATTAAGCAGGGGAGGAGTAAGAAA 2520  
QY 2521 TCCTCGAGATGAGCTGCAGTTTAAATGGGCGCTCACTGGGAATGTGACGTTGAGCAGA 2580  
DB 2521 TCCTCGAGATGAGCTGCAGTTTAAATGGGCGCTCACTGGGAATGTGACGTTGAGCAGA 2580  
QY 2581 GACGTTAGGGAAGTGGATCTTGGCAAGGCAATTCAGGCAGAGAAACAAGATGTGCACTG 2640  
DB 2581 GACGTTAGGGAAGTGGATCTTGGCAAGGCAATTCAGGCAGAGAAACAAGATGTGCACTG 2640  
QY 2641 CCCCAAGTGAAGCTTCTCTAGTGTGAGGAAAGCAGAGGAGACCAAGCAGAGTGC 2700  
DB 2641 CCCCAAGTGAAGCTTCTCTAGTGTGAGGAAAGCAGAGGAGACCAAGCAGAGTGC 2700  
QY 2701 TGGCAGAGGCTAGAAATGGAAGGAGGCGCTGCGGAGGACAGGCTGAGGAGGCTTGG 2760  
DB 2701 TGGCAGAGGCTAGAAATGGAAGGAGGCGCTGCGGAGGACAGGCTGAGGAGGCTTGG 2760  
QY 2761 CTTCTGCTAAGTGAAGTGGGAAACCACTGGAGGGTTTGAACAGAGGAGTGCCTTGAATTGAT 2820  
DB 2761 CTTCTGCTAAGTGAAGTGGGAAACCACTGGAGGGTTTGAACAGAGGAGTGCCTTGAATTGAT 2820  
QY 2821 TTATATTTTGAAGGCTCATTTAGCTGCAATATTTGTGAAAACTTTAGTGGCAAGGGC 2880  
DB 2821 TTATATTTTGAAGGCTCATTTAGCTGCAATATTTGTGAAAACTTTAGTGGCAAGGGC 2880  
QY 2881 AGAAGGAGAGGGAAGACCTGTTAGGAAGCTTCTGCAAGGTTCCAGGCTTGGGCTTGGGC 2940  
DB 2881 AGAAGGAGAGGGAAGACCTGTTAGGAAGCTTCTGCAAGGTTCCAGGCTTGGGCTTGGGC 2940  
QY 2941 CACAGCAACAGCAGTGTGCAATATCTAGATTTATTTTGAAGAGCCCAATAGGATTTGC 3000  
DB 2941 CACAGCAACAGCAGTGTGCAATATCTAGATTTATTTTGAAGAGCCCAATAGGATTTGC 3000  
QY 3001 TGAGAGTTTGAATGTGGAGTGTAAAGAAAGAGAGTTAATGATGACATTAAGGTTTTTG 3060  
DB 3001 TGAGAGTTTGAATGTGGAGTGTAAAGAAAGAGAGTTAATGATGACATTAAGGTTTTTG 3060  
QY 3061 GCCTGAATACAGGAAAGATGAGGTTTACAGTTTCTGAAATAGGGAAGGATGGCTGGGT 3120  
DB 3061 GCCTGAATACAGGAAAGATGAGGTTTACAGTTTCTGAAATAGGGAAGGATGGCTGGGT 3120  
QY 3121 AAGTAAGGAATTTGGTGCAAGCAGGCTGTCTGTGTGGATTTGGGAGGTTCTGGCTGCA 3180



Db 3121 AAGTAGGATTTGGTGGAGAGAGGCTGCTGTGGTGGATGGAGGTTCTGGCTGCA 3180  
Qy 3181 AATCAAGTGGAGATTTCTCTAGGTCAGGTCGTGACAGAGAGCTGACAGAGGATCTGAA 3240  
Db 3181 AATCAAGTGGAGATTTCTCTAGGTCAGGTCGTGACAGAGAGCTGACAGAGGATCTGAA 3240  
Qy 3241 TGCACATGGTTTATTTGTTGGGGTGTCTCTCAGAGGACCTGTGAAAGCCCTTTATCAGTC 3300  
Db 3241 TGCACATGGTTTATTTGTTGGGGTGTCTCTCAGAGGACCTGTGAAAGCCCTTTATCAGTC 3300  
Qy 3301 ATTATTTGGCTGTGAGAAATTTCTCTGGAGTGTGGTACATTTGAAGCAAGTGAATTTCA 3360  
Db 3301 ATTATTTGGCTGTGAGAAATTTCTCTGGAGTGTGGTACATTTGAAGCAAGTGAATTTCA 3360  
Qy 3361 GTTGAGGCAAGTCTCTGGAAGAGAGCTGTAGGATCTGGCAGCTACCATGGCTGGTAG 3420  
Db 3361 GTTGAGGCAAGTCTCTGGAAGAGAGCTGTAGGATCTGGCAGCTACCATGGCTGGTAG 3420  
Qy 3421 TGTGTTGGGGTGGGGTCTCTGGGCACTGGCTGTGTGAAGGATCTGGCAGGACACACA 3480  
Db 3421 TGTGTTGGGGTGGGGTCTCTGGGCACTGGCTGTGTGAAGGATCTGGCAGGACACACA 3480  
Qy 3481 GCGCCCTCTACTCAACATCAGCATCTCAGTGGCATTTAAAGCCATGAGCTGGAGGGC 3540  
Db 3481 GCGCCCTCTACTCAACATCAGCATCTCAGTGGCATTTAAAGCCATGAGCTGGAGGGC 3540  
Qy 3541 CATGAGATGTTCTGTAGTATTACTGAGAAGCAACAGAAAGAGCCATGATGGAGGCC 3600  
Db 3541 CATGAGATGTTCTGTAGTATTACTGAGAAGCAACAGAAAGAGCCATGATGGAGGCC 3600  
Qy 3601 TTGGGCTCTCTGGAAATGGAAATCAGCCAAAGAGCTGAGAAGGAGTTACTTAAAGTTC 3660  
Db 3601 TTGGGCTCTCTGGAAATGGAAATCAGCCAAAGAGCTGAGAAGGAGTTACTTAAAGTTC 3660  
Qy 3661 AGAGAAACACAGAGAGTGTGTCTGAGAGCTGAGTTTCTTTATTTCACTCATTC 3720  
Db 3661 AGAGAAACACAGAGAGTGTGTCTGAGAGCTGAGTTTCTTTATTTCACTCATTC 3720  
Qy 3721 CCTTCTCCAAATTAAGCCACTGTGTAGTTGGGCCCTCCAGGGTTGAAAGGCAAGAGGAGA 3780  
Db 3721 CCTTCTCCAAATTAAGCCACTGTGTAGTTGGGCCCTCCAGGGTTGAAAGGCAAGAGGAGA 3780  
Qy 3781 AAGCAGAGCTTTGGAAACAGAGCTTTTCTGCAATAGCTGGAGAGGATTAAGGA 3840  
Db 3781 AAGCAGAGCTTTGGAAACAGAGCTTTTCTGCAATAGCTGGAGAGGATTAAGGA 3840  
Qy 3841 TAGAGTGTAAATAAAAAAAAAAAAAAAAAAAAA 3871  
Db 3841 TAGAGTGTAAATAAAAAAAAAAAAAAAAAAAAA 3871

## RESULT 3

ABA90876

ID ABA90876 standard; cDNA; 3871 BP.

XX ABA90876;

XX ABA90876;

XX 14-FEB-2002 (first entry)

XX Human polynucleotide SEQ ID NO 1.

XX Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;  
KW immune disorder; bacterial infection; fungal infection; cancer; tumour;  
KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibitor;  
KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;  
KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnery;  
KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;  
KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;  
KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.  
OS Homo sapiens.

XX PN US2001039335-A1.  
XX PD 08-NOV-2001.  
XX PF 04-DEC-2000; 2000US-0729674.  
XX PR 10-APR-1997; 97US-126425P.  
PR 04-DEC-1997; 97US-067454P.  
PR 20-DEC-1997; 97US-068379P.  
PR 02-JAN-1998; 98US-070346P.  
PR 07-JAN-1998; 98US-070643P.  
PR 08-JAN-1998; 98US-070755P.  
PR 13-JAN-1998; 98US-071304P.  
PR 22-JAN-1998; 98US-072134P.  
PR 30-JAN-1998; 98US-073095P.  
PR 18-FEB-1998; 98US-075038P.  
PR 30-MAR-2000; 2000US-0539330.  
PR 23-NOV-1998; 98US-0197886.  
XX (JACO/) JACOBS K.  
PA (MCCO/) MCCOY J M.  
PA (LAVA/) LAVALLIE E R.  
PA (COLL/) COLLINS-RACIE L A.  
PA (EVAN/) EVANS C.  
PA (MERE/) MERBERG D.  
PA (TREA/) TREACY M.  
PA (AGOS/) AGOSTINO M J.  
PA (STEL/) STEININGER R J.  
PA (SPAU/) SPAULDING V.  
PA (WONG/) WONG G G.  
PA (CLAR/) CLARK H.  
PA (FECH/) FECHTEL K.  
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C,  
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;  
PI Weig GG, Clark H, Fechtel K;  
XX WPI; 2002-040725/05.  
DR P-PSDB; ABB55698.  
XX New secreted proteins and encoding polynucleotides, useful in gene  
PT therapies, particularly for preventing or treating autoimmune  
PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,  
PT stroke or inflammations  
XX Claim 1; Page 175-177; 349pp; English.

The invention relates to isolated polynucleotides (ABA90876-ABA90968 and  
ABA90980) and encoded proteins (ABB55698-ABB58000), especially  
polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90985) and  
proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in  
clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1  
are deposited with the American Type Culture Collection (ATCC) with  
accession number 98599. The polynucleotides and encoded polypeptides have  
cytostatic, anti-inflammatory, immunomodulator, vulnery,  
neuroprotective, activin, inhibitor, chemotactic, haemostatic, thrombolytic  
and anti-inflammatory activity and acting as cytokine modulators,  
haematopoiesis regulators, tissue growth modulators and/or cadherin  
suppressors. The polypeptides and polynucleotides are useful in gene  
therapies, particularly for preventing, treating or ameliorating any of  
the following diseases: immune deficiency and disorders; e.g. bacterial  
or fungal infections, autoimmune disorders, cancer, systemic lupus  
erythematosus or graft-versus-host disease; myeloid or lymphoid cell  
deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
osteoarthritis; central and peripheral nervous system diseases and  
neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;  
CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis  
CC or systemic inflammatory response syndrome; ischaemia-reperfusion  
CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or  
CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus  
CC foliaceus.





```
QY 2101 TTCACTTGGCTGACTTGTACCAATTCCTAGCCACCACTGAAACCAAGCTTCACTAGAG 2160
DB 2101 TTCACTTGGCTGACTTGTACCAATTCCTAGCCACCACTGAAACCAAGCTTCACTAGAG 2160
QY 2161 TGTAGAGTCAGAAATGTGCTTTTCCCTCCCTTTGTCATCTCCAAATTTCAACCGGTGG 2220
DB 2161 TGTAGAGTCAGAAATGTGCTTTTCCCTCCCTTTGTCATCTCCAAATTTCAACCGGTGG 2220
QY 2221 CCGATCCCATTTGAGGACCAATGCTAGTTATTAAGTCTCCGAGTTCGAAAGGAGAAAGC 2280
DB 2221 CCGATCCCATTTGAGGACCAATGCTAGTTATTAAGTCTCCGAGTTCGAAAGGAGAAAGC 2280
QY 2281 CAGAGCTGTCTAGTTTCATTCATTCCTTTCAGTAAATTTATTTAGTACCTACTGTGTGC 2340
DB 2281 CAGAGCTGTCTAGTTTCATTCATTCCTTTCAGTAAATTTATTTAGTACCTACTGTGTGC 2340
QY 2341 TAGGCATTCACCTGGGAACTAGAGATCTTACAGAAATAACAGGGAAGTTCCCTGTGCT 2400
DB 2341 TAGGCATTCACCTGGGAACTAGAGATCTTACAGAAATAACAGGGAAGTTCCCTGTGCT 2400
QY 2401 CATGGAGCTTACATTTCTAGGGAGAAAGAGATAGCCATATAGGAATTAATATATAC 2460
DB 2401 CATGGAGCTTACATTTCTAGGGAGAAAGAGATAGCCATATAGGAATTAATATATAC 2460
QY 2461 AAGGTATCATGTAGTATGATTAATCTCTGAGAGAAATTAAGCAGGGGAGGAGTAAGAAA 2520
DB 2461 AAGGTATCATGTAGTATGATTAATCTCTGAGAGAAATTAAGCAGGGGAGGAGTAAGAAA 2520
QY 2521 TCCTGGAGATGAGGCTGCAATTTTAAATGGGCTCCTACCTGGGAATGTGACGTTGAGCAGA 2580
DB 2521 TCCTGGAGATGAGGCTGCAATTTTAAATGGGCTCCTACCTGGGAATGTGACGTTGAGCAGA 2580
QY 2581 GAGCTTAGGGAAGTGGATCTTGACCAAGCATTTCCAGGCAGAGGAACAGATGTGCACTG 2640
DB 2581 GAGCTTAGGGAAGTGGATCTTGACCAAGCATTTCCAGGCAGAGGAACAGATGTGCACTG 2640
QY 2641 CCCAAATGAGAACTTCTCTACGTGTGTCAGAAAGAGCAGGAGACCAAGCAGAGTCG 2700
DB 2641 CCCAAATGAGAACTTCTCTACGTGTGTCAGAAAGAGCAGGAGACCAAGCAGAGTCG 2700
QY 2701 TGGCAGGGGTGAATTTGGAAGAGAGCGGCTGGGAGGACACAGTGTGTGGAGGCGCTTGG 2760
DB 2701 TGGCAGGGGTGAATTTGGAAGAGAGCGGCTGGGAGGACACAGTGTGTGGAGGCGCTTGG 2760
QY 2761 CTTCTGCTAGTGTGAGTGGAACTCTGAGGCTTGAACAGAGAGTGCCTTGAATGAT 2820
DB 2761 CTTCTGCTAGTGTGAGTGGAACTCTGAGGCTTGAACAGAGAGTGCCTTGAATGAT 2820
QY 2821 TTATATTTTCCAAAGGCTCATTTCTAGCTGCAATATTTGTGAAACCTTTAGTGGACAAAGGC 2880
DB 2821 TTATATTTTCCAAAGGCTCATTTCTAGCTGCAATATTTGTGAAACCTTTAGTGGACAAAGGC 2880
QY 2881 AAGAGAGAGAGAGAACCTTTAGAAAGTACTGCAAGGTTCCAGGCTTGGGCTGGGC 2940
DB 2881 AAGAGAGAGAGAGAACCTTTAGAAAGTACTGCAAGGTTCCAGGCTTGGGCTGGGC 2940
QY 2941 CACAGCAACAGCAGTGTCAATATCTAGATTTATTTTGAAGAGCCCAATAGGATTTGC 3000
DB 2941 CACAGCAACAGCAGTGTCAATATCTAGATTTATTTTGAAGAGCCCAATAGGATTTGC 3000
QY 3001 TGAGAGTTTGAATGTGAGTGTGAGAGAGAGAGATTAATGATGACATTAAGGTTTTTG 3060
DB 3001 TGAGAGTTTGAATGTGAGTGTGAGAGAGAGAGATTAATGATGACATTAAGGTTTTTG 3060
QY 3061 GCCTGAATAGCAGAAAGATGAGTACCAGTTACTGAAATAGGGAAGGATGGGCTGGGT 3120
DB 3061 GCCTGAATAGCAGAAAGATGAGTACCAGTTACTGAAATAGGGAAGGATGGGCTGGGT 3120
QY 3121 AAGTAAAGGAATTTGTCAGAGAGAGCTGCTCTGCTGGAATGGAGGTTCTGCTGCA 3180
DB 3121 AAGTAAAGGAATTTGTCAGAGAGAGCTGCTCTGCTGGAATGGAGGTTCTGCTGCA 3180
```

```
QY 3181 AATCAAGTGGAGATTCTCTCAGCTCAGGTCTGAGCAGAGCTCGAGACAGGATCTGAA 3240
DB 3181 AATCAAGTGGAGATTCTCTCAGCTCAGGTCTGAGCAGAGCTCGAGACAGGATCTGAA 3240
QY 3241 TGCATCTGGTTATTTGTTGGGGTCTCTCAGAAAGAACCTGTGAAAGCCTTTATCAGTC 3300
DB 3241 TGCATCTGGTTATTTGTTGGGGTCTCTCAGAAAGAACCTGTGAAAGCCTTTATCAGTC 3300
QY 3301 ATTTATTTGGCTGTGAGAGATTCTCTGGGAGTGTGGGTACATTTGAAGGCAAGTGACTTCA 3360
DB 3301 ATTTATTTGGCTGTGAGAGATTCTCTGGGAGTGTGGGTACATTTGAAGGCAAGTGACTTCA 3360
QY 3361 GTTGAGGGCAAGTCTCTGGAAGAGAGCTGTAGGCATCTGGCAGCTACCATGCTGGTAG 3420
DB 3361 GTTGAGGGCAAGTCTCTGGAAGAGAGCTGTAGGCATCTGGCAGCTACCATGCTGGTAG 3420
QY 3421 TGTGTTGGGGGTGGGGGTCTCTGGGCACTGGCTGTGTGAAGGGATCTGGCAGGGCACCA 3480
DB 3421 TGTGTTGGGGGTGGGGGTCTCTGGGCACTGGCTGTGTGAAGGGATCTGGCAGGGCACCA 3480
QY 3481 GGGCCCCCTACTGAACCATCAGCATGTCTAGTGGCATTTAAAGCCATGCTGGAGGGC 3540
DB 3481 GGGCCCCCTACTGAACCATCAGCATGTCTAGTGGCATTTAAAGCCATGCTGGAGGGC 3540
QY 3541 CACTCAGATTTGCTCTGAGTATTACTGAGAAAGCAACAGAAAGAGCCATGGATGGAGCCC 3600
DB 3541 CACTCAGATTTGCTCTGAGTATTACTGAGAAAGCAACAGAAAGAGCCATGGATGGAGCCC 3600
QY 3601 TTGGGCTCTCTGGGAATGGGAATCAGCCAAAGACTGAGAGGATTAACCTTAGGTC 3660
DB 3601 TTGGGCTCTCTGGGAATGGGAATCAGCCAAAGACTGAGAGGATTAACCTTAGGTC 3660
QY 3661 AGAGAAAAACCAAGAGAGTGTGTCTCTGGAAGCTGAGCTTTCTTTATCAACCTCATTC 3720
DB 3661 AGAGAAAAACCAAGAGAGTGTGTCTCTGGAAGCTGAGCTTTCTTTATCAACCTCATTC 3720
QY 3721 CCTTCTCAATTAAGCCACTTGTCTGCTGGCCCTCCAGGCTTGAGGCAAGAGAGA 3780
DB 3721 CCTTCTCAATTAAGCCACTTGTCTGCTGGCCCTCCAGGCTTGAGGCAAGAGAGA 3780
QY 3781 AAGGCACAGCGTTTGGGAAACAAGACTTTCTGCAATAGCCTGGGAAGGAATAAAGGA 3840
DB 3781 AAGGCACAGCGTTTGGGAAACAAGACTTTCTGCAATAGCCTGGGAAGGAATAAAGGA 3840
QY 3841 TAGAGTGTATAATAAATAAATAAATAAATAA 3871
DB 3841 TAGAGTGTATAATAAATAAATAAATAAATAA 3871
```

## RESULT 4

AAF16225 standard; cDNA; 2549 BP.

AC AAF16225;

XX AAF16225;

XX 13-MAR-2001 (first entry)

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:660.

Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;  
vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;  
antibacterial; gene therapy; neural; immune; reproductive; renal;  
gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
wound; infectious disease; ss.

OS Homo sapiens.

XX WO200055174-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05988.

XX 12-MAR-1999; 99US-0124270.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX (ROSE/) ROSEN C A.  
 XX Rosen CA, Ruben SM;  
 XX WPI; 2000-587513/55.  
 XX P-P8DB; AAB57022.  
 XX  
 XX Prostate cancer associated gene sequences, referred to as prostate  
 XX cancer antigens, useful for treatment, prevention, and diagnosis of  
 XX disorders such as prostate cancer -  
 XX  
 XX Claim 1; Page 1097-1098; 238pp; English.  
 XX  
 XX AAF1566 to AAF16505 encode the human prostate cancer associated  
 XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
 XX The prostate cancer antigens can have neuroprotective, cytostatic,  
 XX cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,  
 XX nephrotropic, antiinfective, gynecological and antibacterial activities,  
 XX and can be used in gene therapy. The prostate cancer antigen  
 XX polynucleotides may be used for detection of prostate cancer, chromosome  
 XX identification, as chromosome markers, and for numerous other diagnostic  
 XX or research purposes. The prostate cancer antigens may be used to treat  
 XX disorders such as neural, immune, muscular, reproductive,  
 XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 XX disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
 XX AAB57303 represent sequences used in the exemplification of the present  
 XX invention.  
 XX  
 XX Sequence 2549 BP; 667 A; 519 C; 683 G; 673 T; 7 other;  
 XX  
 XX Query Match 53.0%; Score 2051.2; DB 21; Length 2549;  
 XX Best Local Similarity 99.2%; Pred. No. 0;  
 XX Matches 2077; Conservative 6; Mismatches 4; Indels 7; Gaps 2;  
 XX  
 XX 1763 CATGCCGGCCCTTCTCACCCTTACACCTGCTCTTATCTCCTCACAATCTGTTTCA 1822  
 XX 206 CTGCGCGGGCCCTTCTCACCCTTACACCTGCTCTTATCTCCTCACAATCTGTTTCA 265  
 XX 1823 CTTTCATCCCTGCTCTCTCTCCTCATGTTTACACCTGCTCTTCCCATGTTTCCCATGCTTTC 1882  
 XX 266 CTTTCATCCCTGCTCTCTCTCCTCATGTTTACACCTGCTCTTCCCATGTTTCCCATGCTTTC 325  
 XX 1893 TTACCAATTTGTTTGAAGGCGAGTCTTCTGCTGCTGTTTCTGTTTCTGTTTCTGTTTCTGTTT 1942  
 XX 326 TTACCAATTTGTTTGAAGGCGAGTCTTCTGCTGCTGTTTCTGTTTCTGTTTCTGTTTCTGTTT 385  
 XX 1943 TCAGTATTATTTTAAATAAGAAAAACATTCCTAGAGATGATAATTTGTAAAAACCTCC 2002  
 XX 386 TCAGTATTATTTTAAATAAGAAAAACATTCCTAGAGATGATAATTTGTAAAAACCTCC 445  
 XX 2003 TTGCTCTATTGCTTTTCCAGATTTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2062  
 XX 446 TTGCTCTATTGCTTTTCCAGATTTTGTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 505  
 XX 2063 GAAGACATAGGCTAAATTTCTCCAGCTTCAATAGTCTTCACTTGTGCTGCTGCTGCTGCTGCTGCT 2122  
 XX 506 GAAGACATAGGCTAAATTTCTCCAGCTTCAATAGTCTTCACTTGTGCTGCTGCTGCTGCTGCTGCT 565  
 XX 2123 AATTTAGCACCCTAGTAAACAGTTGAGTACAGAGTGTAGAGTGTAGAGTGTAGAGTGTAGAGTGTAG 2182  
 XX 566 AATTTAGCACCCTAGTAAACAGTTGAGTACAGAGTGTAGAGTGTAGAGTGTAGAGTGTAGAGTGTAG 625  
 XX 2183 TTGCCCCCATTTGTCATCTCCAAATTTACAAAGGTTGGCCGATCCCATTTTGGAGACAATG 2242  
 XX 626 TTGCCCCCATTTGTCATCTCCAAATTTACAAAGGTTGGCCGATCCCATTTTGGAGACAATG 685  
 XX 2243 CTTAGTTATAGTCTCCAGTTGGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2302  
 XX 686 CTTAGTTATAGTCTCCAGTTGGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 745

2303 TTCTTTCAAGTAAATATTTTATGAGTACCTACTGTGTGCTAGGCAATTCACCTGG-----G 2356  
 746 TTCTTTCAAGTAAATATTTTATGAGTACCTACTGTGTGCTAGGCAATTCACCTGGGAACTAG 805  
 2357 AACTAGAGATACCTTCACAGAAATAACAGGAAAGTTCCCTGTGTCTCATGAGCTTACATTC 2416  
 806 AACTAGAGATACCTTCACAGAAATAACAGGAAAGTTCCCTGTGTCTCATGAGCTTACATTC 865  
 2417 TACAGGGAGAAAGAGATAGCAATACATAGGAATAAATATATATACAAAGGTATCATGTAGTGT 2476  
 866 TACAGGGAGAAAGAGATAGCAATACATAGGAATAAATATATATACAAAGGTATCATGTAGTGT 925  
 2477 AATATTGCTGTGGAGAAATAAAGCAGGAGAGGAGTAAAGAAATCCTGGAGATGAGGCT 2536  
 936 AATATTGCTGTGGAGAAATAAAGCAGGAGAGGAGTAAAGAAATCCTGGAGATGAGGCT 985  
 2537 GCAGTTTTAAATGGGGCCTCCTCCTGGGAATGTGACGTTGAGCAGAGACGTTAGGGAAGTGG 2596  
 986 GCAGTTTTAAATGGGGCCTCCTCCTGGGAATGTGACGTTGAGCAGAGACGTTAGGGAAGTGG 1045  
 2597 ATCTGCAAGAGGATTCAGGAGAGAGAAACAAGATGTGCACTGCCCCCAAGTGAAGT 2656  
 1046 ATCTGCAAGAGGATTCAGGAGAGAGAAACAAGATGTGCACTGCCCCCAAGTGAAGT 1105  
 2657 TGCTCTACGTTGTCAGGAAAGAGCAGGAGACCAAGCAGAGTCTGCGCAGGAGGTAGAAT 2716  
 1106 TGCTCTACGTTGTCAGGAAAGAGCAGGAGACCAAGCAGAGTCTGCGCAGGAGGTAGAAT 1165  
 2717 GGAAGGAGAGGCGCTGGGAGAGACAGTGTGTGGAGGCGCTTGGCTTCTGCTAAGTGAGA 2776  
 1166 GGAAGGAGAGGCGCTGGGAGAGACAGTGTGTGGAGGCGCTTGGCTTCTGCTAAGTGAGA 1225  
 2777 TGGGAACCACTGGAGGCTTTTGAACAGAGGAGTGCCTTTGATTGATTATATTTTGAAGGG 2836  
 1226 TGGGAACCACTGGAGGCTTTTGAACAGAGGAGTGCCTTTGATTGATTATATTTTGAAGGG 1285  
 2837 TCATCTAGTGCATATTTGTGAATAACCTTGTGACAGAGGCGCAGAGAGAGAGAGAG 2896  
 1286 TCATCTAGTGCATATTTGTGAATAACCTTGTGACAGAGGCGCAGAGAGAGAGAGAGAG 1345  
 2897 ACCTGTTAGGAAGCTACTGCAAGGTTCCAGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 2956  
 1346 ACCTGTTAGGAAGCTACTGCAAGGTTCCAGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT 1405  
 2957 GTCAATATCTAGATTTATTTTGAAGAGCCATAGGATTTGCTGAGAGTTTGAATGTG 3016  
 1406 GTCAATATCTAGATTTATTTTGAAGAGCCATAGGATTTGCTGAGAGTTTGAATGTG 1465  
 3017 GAGTGTAGAGAGAGAGAGTAAATGATGACATTAAGGTTTTTTGGCCTGAAATAGCAGGAA 3076  
 1466 GAGTGTAGAGAGAGAGAGTAAATGATGACATTAAGGTTTTTTGGCCTGAAATAGCAGGAA 1525  
 3077 AGATGAGATTTACAGTTTACTGAAATAGGAGAGGAGGCTGGGTGAGTAAAGAAATTTGGT 3136  
 1526 AGATGAGATTTACAGTTTACTGAAATAGGAGAGGAGGCTGGGTGAGTAAAGAAATTTGGT 1585  
 3137 GCAAGCAGGCTGTCTGTGTTGGAAATGGAGAGGTTCTGGCTGCAAAATCAAAAGTGGAGA-T 3195  
 1586 GCAAGCAGGCTGTCTGTGTTGGAAATGGAGAGGTTCTGGCTGCAAAATCAAAAGTGGAGAGT 1645  
 3196 TCTCTCAGTTCAGGCTGTGAGAGAGAGCTGAGACAGGAGTCTGAATGCACTTGGTTTATT 3255  
 1646 TCTCTCAGTTCAGGCTGTGAGAGAGAGCTGAGACAGGAGTCTGAATGCACTTGGTTTATT 1705  
 3256 GTTGGGGGTGCTCTCAGAGAGAAACCTGTGTGAAAGCCTTTATCAGTCAATTTATTGGCTGTGA 3315  
 1706 GTTGGGGGTGCTCTCAGAGAGAAACCTGTGTGAAAGCCTTTATCAGTCAATTTATTGGCTGTGA 1765  
 3316 GAGATTTCTCTGGAGTGTGGTACATTTGAAGCAGAGTCACTTCAAGTTCAGTTCAGTTCAGTTCAGT 3375  
 1766 GAGATTTCTCTGGAGTGTGGTACATTTGAAGCAGAGTCACTTCAAGTTCAGTTCAGTTCAGTTCAGT 1825

3376	QT	CTGGAAAGAGCGCTGTAGGCATCTGGCAGCTACCATGCGTGTAGTGTGTGGGGGTGGG	3433
1826	Db	CTGGAAAGAGCGCTGTAGGCATCTGGCAGCTACCATGCACTGGTGTGTGGGGGTGGG	1885
3436	QY	GCTCTCTGGGCACTGGCTGTGTGAAGGGAICTGGCAGGGCACCACAGCGCCCCCTACTGAA	3495
1886	Db	GCTCTCTGGGCACTGGCTGTGTGAAGGGAICTGGCAGGGCACCACAGCGCCCCCTACTGAA	1945
3496	QY	CCATCAGCATCTCAGTGGCATTTAAAGCCATGCACTGGAGGGGCACTGAGATTGTCTC	3555
1946	Db	CCATCAGCATCTCAGTGGCATTTAAAGCCATGCACTGGAGGGGCACTGAGATTGTCTC	2005
3556	QY	TGAGTATTACTGAGAAGCAACAGAAAGAGCCATGATGGAGCCCTTGGGCTCTCTGGGA	3615
2006	Db	TGAGTATTACTGAGAAGCAACAGAAAGAGCCATGATGGAGCCCTTGGGCTCTCTGGGA	2065
3616	QY	AATGGGAATCAGCCAAAGAGCTCGAAAGAGGTACCTTAAGGTGAGAGAAACCAAGAG	3675
2066	Db	AATGGGAATCAGCCAAAGAGCTCGAAAGAGGTACCTTAAGGTGAGAGAAACCAAGAG	2125
3676	QY	AGTGTGTGTCTCGAAGCTGAGCTTCTTTATTCAACCTCATTCCTTCTCCAAATAAG	3735
2126	Db	AGTGTGTGTCTCGAAGCTGAGCTTCTTTATTCAACCTCATTCCTTCTCCAAATAAG	2185
3736	QY	CCACTTGTGTAGTTGGCCCTCCAGGGTTGAAGCAGAGGAGAAAGGCACAGCGTTTG	3795
2186	Db	CCACTTGTGTAGTTGGCCCTCCAGGGTTGAAGCAGAGGAGAAAGGCACAGCGTTTG	2245
3796	QY	GGAAACAAGACTTCTCTGCAATAGCTGGGAAGGAATAAAGGATAGAGTGT 3849	
2246	Db	GGAAACAAGACTTCTCTGCAATAGCTGGGAAGGAATAAAGGATAGAGTGT 2299	

RESULT 5  
 AAH14525  
 ID AAH14525 standard, cDNA, 1961 BP.  
 XX  
 AC AAH14525;  
 XX  
 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:12066.  
 XX  
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 EP1074617-A2.  
 XX  
 07-FEB-2001.  
 XX  
 28-JUL-2000; 2000EP-0116126.  
 XX  
 29-JUL-1999; 99JP-0248036.  
 PR  
 27-AUG-1999; 99JP-0300253.  
 PR  
 11-JAN-2000; 2000JP-0118776.  
 PR  
 02-MAY-2000; 2000JP-0183767.  
 PR  
 09-JUN-2000; 2000JP-0241899.  
 XX  
 (HELI-) HELIX RES INST.  
 PA  
 Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 WPI; 2001-318749/34.  
 DR  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 XX full-length cDNAs -  
 XX  
 Claim 8; SEQ ID 12066; 2537pp + CD ROM; English.  
 XX  
 XX

584	QY	ASCTGTTTCATCCGAGCGATGCGGGCATACAACGGTGAGAACTGAGAACATCCATCAC	743
661	Db	ASCTCGTTTCATCCGAGCAGTGGGGCATACAACGGTGAGAACTGAGAACATCCATCAC	720
744	QY	GACATGAGAGCTGGCCCTTCCGACTTCTTCAAGAGCCCTTTACGAGTGTCTCGCAGCCTGC	803
721	Db	GACATGGAGCTGGCCCTTCCGACTTCTTCAAGAGCCCTTTACGAGTGTCTCGCAGCCTGC	780
804	QY	GAGGTTCCAGGAGATCAAGGATTTCAAGGATTTCTACCTTTCCATAGCAGATCATAT	863
781	Db	GAGGTTCCAGGAGATCAAGGATTTCAAGGATTTCTACCTTTCCATAGCAGATCATAT	840
864	QY	GTAGAAGTTCTGGAATGCAAAATACACAGTGTGAAGAGAACTCACCCCAAGTTATAGGAGGC	923
841	Db	GTAGAAGTTCTGGAATGCAAAATACACAGTGTGAAGAGAACTCACCCCAAGTTATAGGAGGC	900
924	QY	TATCCGGTTGAGAAATTTGGCTACCATGTATCATTTCTGCAAGTTTGCCTATTATAAG	983
901	Db	TATCCGGTTGAGAAATTTGGCTACCATGTATCATTTCTGCAAGTTTGCCTATTATAAG	960
984	QY	TTGAACACCTGGAAGATTCAGACCCCTGTGCAGTCAGCTATCTCTCTTTTGATCAGAAT	1043
961	Db	TTGAACACCTGGAAGATTCAGACCCCTGTGCAGTCAGCTATCTCTCTTTTGATCAGAAT	1020
1044	QY	GACAGGTTCATGCAGCAGAACCTTGGTGTAATACAGGTACACAGAGACACTTGGGGCTC	1103
1021	Db	GACAGGTTCATGCAGCAGAACCTTGGTGTAATACAGGTACACAGAGACACTTGGGGCTC	1080
1104	QY	TCGGATGAGCACTTCCAGCCGACACCTGGAAGCAGTTTCAGTTCTTTTAATGTGACCACACTC	1163
1081	Db	TCAGATGAGCACTTCCAGCCGACACCTGGAAGCAGTTTCAGTTCTTTTAATGTGACCACACTC	1140
1164	QY	CAGAAGGAGCTGTATGACTTTGTCTGAGGAAAAATATAATGGATGATGATGAGGGAAGATT	1223
1141	Db	CAGAAGGAGCTGTATGACTTTGTCTGAGGAAAAATATAATGGATGATGATGAGGGAAGATT	1200
1224	QY	GTGGAATATGTGGATGACTCTTTGGAACTGGAGAGACCGAGCTAGCCACAGAACCAAA	1283
1201	Db	GTGGAATATGTGGATGACTCTTTGGAACTGGAGAGACCGAGCTAGCCACAGAACCAAA	1260
1284	QY	GAGACTTCTCTTGGCGTTCAGAAACACAGATTCTTTCTCTTTTCCCAACAGCCAGG	1343
1261	Db	GAGACTTCTCTTGGCGTTCAGAAACACAGATTCTTTCTCTTTTCCCAACAGCCAGG	1320
1344	QY	CTGTTGATACCTCAGAGCTTCTTTTACTCTTCAAAGTGAAGGGAAGCCCGTCTCT	1403
1321	Db	CTGTTGATACCTCAGAGCTTCTTTTACTCTTCAAAGTGAAGGGAAGCCCGTCTCT	1380
1404	QY	CTAACTGATGTCACTCAGGGGTGAGCGCTTCTATCTTTCACACCTGCCACTCATG	1463
1381	Db	CTAACTGATGTCACTCAGGGGTGAGCGCTTCTATCTTTCACACCTGCCACTCATG	1440
1464	QY	TTCAACACTATCTTCTCACCTTTTTTTTGAGATGAGATCTGCTCTCTTGGCCAGGCTG	1523
1441	Db	TTCAACACTATCTTCTCACCTTTTTTTTGAGATGAGATCTGCTCTCTTGGCCAGGCTG	1500
1524	QY	GAGTGCATGTGCACTCAGCTCACTGCAACCTCCGCTCTTGGGTTCAAGCAATTCT	1583
1501	Db	GAGTGCATGTGCACTCAGCTCACTGCAACCTCCGCTCTTGGGTTCAAGCAATTCT	1560
1584	QY	GCTGCATCAGCCTCCGAGTACTGCGGATTCAGGCAATGTGCCACACGCGCGCTAATT	1643
1561	Db	GCTGCATCAGCCTCCGAGTACTGCGGATTCAGGCAATGTGCCACACGCGCGCTAATT	1620
1644	QY	TTGTATTTTTAGTAGAGACGGGGTTTTGCCATGTTGGCCAGGCTGGTCTCGAACTCTTGA	1703
1621	Db	TTGTATTTTTAGTAGAGACGGGGTTTTGCCATGTTGGCCAGGCTGGTCTCGAACTCTTGA	1680
1704	QY	CTTCAGATGATCCCATCTGCTTGGCTCCCAACAGTCTCGGATTCAGGCGTGAAGCAAC	1763
1681	Db	CTTCAGATGATCCCATCTGCTTGGCTCCCAACAGTCTCGGATTCAGGCGTGAAGCAAC	1740
1764	QY	ATGCCCGCCCTCTTCTCACCTTTTACACCTGTCTTCTTATCTCTCATCTGTTTTTCAC	1823

Db	1741	ATGCGCGGCTCTTCTTCACCTTTACACCTGCTCTTCTTATCTCCTCACATCTGTTTTCACAC	1800
Qy	1824	CTTCATCCCTGTCTTCTCATGTTCACACTGTCTTCTTCCCATGTTTCATAGCTGCCTTTCT	1883
Db	1801	CTTCATCCCTGTCTTCTCATGTTCACACTGTCTTCTTCCCATGTTTCATAGCTGCCTTTCT	1860
Qy	1884	TACCATTTTGGTTGAAGGCGACTCTCTCTCGCTGTGTTTTTGTCTTCCGAGAAAT	1943
Db	1861	TACCATTTTGGTTGAAGGCGACTCTCTCTCGCTGTGTTTTTGTCTTCCGAGAAAT	1920
Qy	1944	CAGTATTATTTTTTAATAGAAAAACATCTCTAGAAGATG	1984
Db	1921	CAGTATTATTTTTTAATAGAAAAACATCTCTAGAAGATG	1961
RESULT 6			
AAH31138	ID	AAH31138 standard; cDNA; 1586 BP.	
XX	AC	AAH31138;	
DT	DT	27-JUL-2001 (first entry)	
XX	XX	Human colon cancer cell line Km12L4-A cDNA library derived seq#1072.	
XX	XX	Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;	
KW	KW	detection; colon cancer cell line Km12L4-A; ss.	
XX	OS	Homo sapiens.	
XX	WO	WO20001891-6-A2.	
PN	PN	06-APR-2000.	
PD	PD		
XX	XX	23-SEP-1999; 99WO-US22226.	
XX	XX	28-SEP-1998; 98US-0102161.	
PR	PR	28-SEP-1998; 98US-0102180.	
PR	PR	29-SEP-1998; 98US-0102380.	
PR	PR	08-OCT-1998; 98US-0103815.	
PR	PR	27-OCT-1998; 98US-0105877.	
XX	XX	(CHIR ) CHIRON CORP.	
PA	PA	(HYSE-) HYSEQ INC.	
XX	XX	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;	
PI	PI	Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;	
PI	PI	Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;	
PI	PI	Leeshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;	
XX	XX	WPI; 2000-293155/25.	
DR	DR		
XX	XX	Polynucleotide library comprising 1079 defined sequences, useful in	
PT	PT	the form of an array to detect cancer or susceptibility to cancer -	
XX	XX	Claim 1; page 498-499; 502pp; English.	
XX	XX		
CC	CC	The present invention describes a library of polynucleotides comprising	
CC	CC	1079 nucleotide sequences (given in AAH31067 to AAH31145). Also described	
CC	CC	are: (1) an isolated polynucleotide (i) having at least 90% identity to	
CC	CC	one of the 1079 sequences; (2) a recombinant host cell containing (1);	
CC	CC	(3) an isolated polypeptide (ii) encoded by (1); (4) an antibody that	
CC	CC	specifically binds to (ii); (5) a vector comprising (1); and (6) a method	
CC	CC	of detecting differentially expressed genes correlated with a cancerous	
CC	CC	state of a mammalian cell comprising detecting a gene product encoded by	
CC	CC	one of the 1079 sequences given in the specification. The polynucleotides	
CC	CC	are used to monitor patients having (or susceptible) to cancer to detect	
CC	CC	potentially malignant events at a molecular level before they are	
CC	CC	detectable at a gross morphological level. The polynucleotides are also	
CC	CC	useful for monitoring the efficacy of various therapies and preventive	
CC	CC	interventions. Polynucleotide probes based on the disclosed sequences	
CC	CC	are useful for chromosome mapping and detection of transcription levels.	

XX









of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH185893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 818 BP; 159 A; 271 C; 236 G; 148 T; 4 other;

Query Match 18.0%; Score 695.6; DB 22; Length 818;  
Best Local Similarity 96.7%; Pred. No. 1.9e-130;  
Matches 739; Conservative 2; Mismatches 18; Indels 5; Gaps 3;

Qy	24	CTTCTTCGTCGCTTCCTTCCTTCCTTCGCGCGGCGCGATGAGCGCGGCGCGCGGGG	83
Db	1	CTTCTTCGTCGCTTCCTTCCTTCCTTCGCGCGGCGCGATGAGCGCGGCGCGGGG	60
Qy	84	GCGCGGCGCTGCTAGCGCTGTGTGCGCTTCGCGCTCGCGCGCGCGCGCGCCAA	143
Db	61	GCGCGGCGCTGCTAGCGCTGTGTGCGCTTCGCGCTCGCGCGCGCGCGCCAA	120
Qy	144	TACGAAGCTTACAGCTTCGCGAGCTTCCACGGGACGAGCTGATGCGCTCGAGTCGGCC	203
Db	121	TACGAAGCTTACAGCTTCGCGAGCTTCCACGGGACGAGCTGATGCGCTCGAGTCGGCC	180
Qy	204	TACCGGACGCGTGTGGAAGAATACAGCGCGGAGCACTGGCGCGAGAGSTKSGCTACCTG	263
Db	181	TACCGGACGCGTGTGGCAAGTACAGCGCGGAGCACTGGCGCGAGAGCGTGGGTACTCTG	240
Qy	264	GAGATCAGCTCGGGCTGCACCGCTTGCTGCGGACAGGAGGCTTCTGCGCACCGCAAC	323
Db	241	GAGATCAGCTCGGGCTGCACCGCTTGCTGCGGACAGGAGGCTTCTGCGCACCGCAAC	300
Qy	324	TGAGCGCGCGCGGAGCGCGGAGCGCGCGCGCTCGCGAAGTATCCCGAGCTCGCG	383
Db	301	TGAGCGCGCGCGGAGCGCGGAGCGCGCGCGCTCGCGAAGTATCCCGAGCTCGCG	360
Qy	384	CTCTTCGGGGGCTGTGCTGGCGCGCGCACTGCCCTCAAGCGCTGCAAGCGGSGCTGCCA	443
Db	361	CTCTTCGGGGGCTGTGCTGGCGCGCGCACTGCCCTCAAGCGCTGCAAGCGGSGCTGCCA	420
Qy	444	GCCTTCGGCGAGTCCAGGCCGAGCGGAGGTGCTGGCGGACTTTCAGCGCGCGGAGGCC	503
Db	421	GCCTTCGGCGAGTCCAGGCCGAGCGGAGGTGCTGGCGGACTTTCAGCGCGCGGAGGCC	480
Qy	504	TACAAGTTCCTGCGAGTTCGCTTACTTCAAGGCAATAATCTCCCGAAAGCCATCGCGCT	563
Db	481	TACAAGTTCCTGCGAGTTCGCTTACTTCAAGGCAATAATCTCCCGAAAGCCATCGCGCT	540
Qy	564	GCTTCACACCTTTCTACTGAGATCCTGATGACGAATGATGAAGAGAAATCGGCTAT	623
Db	541	GCTTCACACCTTTCTACTGAGATCCTGATGACGAATGATGAAGAGAAATCGGCTAT	600
Qy	624	TATAAGAGCTGCTCGGTGCCGAGGACTACATTTAAGACCTGGAAACCAAGT--CATATG	681
Db	601	TATAAGAGCTG--CTGGTCCGAGGACTACATTTAAGACCTGGAAACCAAGTTCATATGA	659
Qy	682	AAGGCTGTTCATCCGAGAGGTGGGGGATCAACGGGTGAGNATGGGAGACATCCATCA	741
Db	660	AAGGCTGTTCATCCGAGNCGTGGGGGATCAACGGGTGAGNATGGGAGACATTCATCA	719
Qy	742	CAGACATGAGCTGGGCC--TTCCCGAGCTTCTTCAAGGCGCTTTT	783

2879 GCAGAGGAGGAGGAGAGCTGTTAGAGCTACTCTCAAGGTTCCAGGCTTGGGCTCG 2938  
182 GCAGAGGAGGAGGAGAGCTGTTAGAGCTACTCTCAAGGTTCCAGGCTTGGGCTCG 241  
2939 GCACAGCAACAGAGGCTGCTCAATATCTAGATTTATTTTGAAGAGCCATAGGATTT 2998  
242 GCACAGCAACAGAGGCTGCTCAATATCTAGATTTATTTTGAAGAGCCATAGGATTT 301  
2999 GCTGAGAGTTGAATGTTGAGTCTAAGAGAGCAAGTAAATGATGACATTAAGGTTT 3058  
302 GCTGAGAGTTGAATGTTGAGTCTAAGAGAGCAAGTAAATGATGACATTAAGGTTT 361  
3059 TGCCCTGAATPAGCAGGAAAGATGAGTTACCACTTACTCAATAGGAAGGATGGCTGG 3118  
362 TGCCCTGAATPAGCAGGAAAGATGAGTTACCACTTACTCAATAGGAAGGATGGCTGG 421  
3119 GTAAGTAAGGAATTTGGTGCAGAGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3178  
422 GTAAGTAAGGAATTTGGTGCAGAGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
3179 CAAATCAAGTGGAGA-TTCTCTCAGCTCAGGCTGCTGAGCAGAGCTCGAGCAGGATCT 3237  
481 CAAATCAAGTGGAGA-TTCTCTCAGCTCAGGCTGCTGAGCAGAGCTCGAGCAGGATCT 540  
3238 GAATGCACTTGGTTATTTGTTGGGGTGTCTCTCAGAGGAACCTGTGAAAGCCTTTATCA 3297  
541 GAATGCACTTGGTTATTTGTTGGGGTGTCTCTCAGAGGAACCTGTGAAAGCCTTTATCA 600  
3298 GTCAATTTATGCTGTGAGAGTTCTCTGGGAGTGGGTACAT 3341  
601 GTCAATTTAT-TCTGTGANAAGTTCTCTGGAAATGGGGTACCT 643

## RESULT 11

AAZ80131/c  
ID AAZ80131 standard; cdna; 618 bp.

XX AC AAZ80131;

XX DT 07-APR-2000 (first entry)

XX DE Human colon cancer cell line SW480 cdna clone SEQ ID NO:215.

XX KW Human; gene expression product; diagnosis; tumour; colon cancer;  
XX KW colorectal adenocarcinoma; cell line SW480; cell proliferation;  
XX KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;

XX OS Homo sapiens.

XX PN WO9964576-A2.

XX PD 16-DEC-1999.

XX PF 09-JUN-1999; 99WO-IB01062.

XX PR 10-JUN-1998; 98US-0088801.

XX PA (FARB) BAYER CORP.

XX PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;  
XX PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;  
XX PI Schlegel R;

XX DR WPI; 2000-087220/07.

XX KW Novel nucleic acids, used to develop products for the diagnosis and  
XX PT treatment of disorders involving unwanted cell proliferation,  
XX PT particularly cancers, especially colon cancer

XX PS Claim 15; Page 218; 469pp; English.

XX XX

CC AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from  
CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
CC cDNA clones can be used to generate antisense oligonucleotides which  
CC can be used for antisense therapy. Methods and products from the present  
CC invention can be used for identifying and/or classifying cancerous cells  
CC present in a human tumour, particularly in solid tumours, e.g.  
CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones  
CC can be used for developing agents for the diagnosis and treatment of  
CC disorders involving unwanted cell proliferation, such as neoplasia,  
CC dysplasia or hyperplasia.

XX SQ Sequence 618 bp; 157 A; 133 C; 169 G; 136 T; 23 other;

Query March 12.08; Score 466.4; DB 21; Length 618;

Best local Similarity 94.8%; Pred. No. 2.3e-84;

Matches 506; Conservative 0; Mismatches 25; Indels 3; Gaps 3;

QY 1076 CCAGTACACAGGACACTTGGGGCTCTCGGATGAGCACTCCAGCCAGCCTGAAGC 1135

DB 534 CNAGTCCCNAGNCCANTGGGGCTTNGAATGAGCANTCCAGCCAGNCTTAAAC 475

QY 1136 AGTTCAGTTCTTTAATGTGACCA-CACTCCAGAAAGGAGC-TGTATGACTTT-GCTAAGGA 1192

DB 474 AGTTCAGTTCTTTAATGTGACCACTCCAGAAAGGAGCTTGTATGACTTTGGNTRAGGA 415

QY 1193 AAATATAATGATGATGAGGAGAGCTTGTGGAATATGTGGATGACCTCTTGAACCT 1252

DB 414 AAATATAATGATGATGAGGAGAGCTTGTGGAATATGTGGATGACCTCTTGAACCT 355

QY 1253 GGAGGAGACAGCTAGCCCAACCAAGAGACTTCTCTTGGCGTTTCAGGAACAC 1312

DB 354 GGAGGAGACAGCTAGCCCAACCAAGAGACTTCTCTTGGCGTTTCAGGAACAC 295

QY 1313 AGATCTTTGTCCTTTTCCCAACAGCCAGGCTGTGATACCTCAGAGCCTTCTCTTTAC 1372

DB 294 AGATCTTTGTCCTTTTCCCAACAGCCAGGCTGTGATACCTCAGAGCCTTCTCTTTAC 235

QY 1373 TCTCCAAAGTGAAGGAGAGCCCGTCTCTCTAACTGATGTCTATCAGGGGTGACCTG 1432

DB 234 TCTCCAAAGTGAAGGAGAGCCCGTCTCTCTAACTGATGTCTATCAGGGGTGACCTG 175

QY 1433 CTTTCTCTATCTCACACCTGCACTCATGTTCACACCTATCTTCTCACCCTTTTTT 1492

DB 174 CTTTCTCTATCTCACACCTGCACTCATGTTCACACCTATCTTCTCACCCTTTTTT 115

QY 1493 GAGATGGAGTCTCGCTCTCTTCCAGGCTGAGTGCATGCAATGCGAGTCTCAGCTCAGCTG 1552

DB 114 GAGATGGAGTCTCGCTCTCTTCCAGGCTGAGTGCATGCAATGCGAGTCTCAGCTCAGCTG 55

QY 1553 CAACCTCCGCTCTTGGGTTCAAGCAATTTCTGCTGATCAGCTCCCGAGTACC 1606

DB 54 CAACCTCCGCTCTTGGGTTCAAGCAATTTCTGCTGATCAGCTCCCGAGTACC 1

## RESULT 12

ABL67662

ID ABL67662 standard; DNA; 2347 bp.

XX AC ABL67662;

XX DT 15-MAY-2002 (first entry)

XX DE Oesophagus cancer related gene sequence SEQ ID NO:5999.

XX KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
XX KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
XX KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;  
XX KW gene; ds.

XX OS Homo sapiens.

XX PN WO200194629-A2.

XX XX



901 ACTACCTGCAGTTTGGCTACTATAGTTGATGATGTCGCCAGGCTGCCGAGCGCG 960  
1018 TCAGCTATCTGCTCTTTGATCAGATGACAAAGTCTATGAGCAGAACCTGGTATTACC 1077  
961 CCAGCTACATGCTCTTTCGACCCCAAGGACAGCGCTCATGCGAGCAGAACCTGGTATTACC 1020  
1078 AGTACCACAGGACACTTTGGGGCTCTCGGATGAGCAGCTCCAGCCCGACCTGAGCAG 1137  
1021 GGTTCACCGGGCTGCTGGGCTGAGAGAGGAGACTTCAGCCCCGGAGGAGGCA 1080  
1138 TTGAGTCTTTAATGAGCAGCAGCTCCAGAGGAGCTGTATGACTTTGTAAGGAAATA 1197  
1081 TGCTCTACCAACACAGACCGCCGAGCTGCGGAGCTGCTGGAGTTTCAACCCATGTACC 1140  
1198 TAATGATGATGATGAGGAGAGTGTGGA 1228  
1141 TGCAGTCAGATGATGAGAGCTGGAGGA 1171

## RESULT 13

AAH30585  
ID AAH30585 standard; cDNA; 408 BP.

AC AAH30585;

XX 27-JUL-2001 (first entry)

XX Human colon cancer cell line Kml2L4-A cDNA library derived sequence #519.

XX Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;  
XX detection; colon cancer cell line Kml2L4-A; ss.

XX Homo sapiens.

XX WO200018916-A2.

XX 06-APR-2000.

XX 23-SEP-1999; 99WO-US222226.

XX 28-SEP-1998; 98US-0102161.

XX 28-SEP-1998; 98US-0102180.

XX 29-SEP-1998; 98US-0102380.

XX 08-OCT-1998; 98US-0103815.

XX 27-OCT-1998; 98US-0105877.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis WA, Garcia PD, Sudduth-Klinger J;  
XX Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
XX Lameon G, Drmanac R, Ctkvenjakov R, Dickson M, Drmanac S, Labat I;  
XX Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
XX WPI; 2000-293155/25.

XX Polynucleotide library comprising 1079 defined sequences, useful in  
XX the form of an array to detect cancer or susceptibility to cancer -

XX Claim 1; Page 333; 502pp; English.

XX The present invention describes a library of polynucleotides comprising  
XX 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described  
XX are: (1) an isolated polynucleotide (I) having at least 90% identity to  
XX one of the 1079 sequences; (2) a recombinant host cell containing (I);  
XX (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that  
XX specifically binds to (II); (5) a vector comprising (I); and (6) a method  
XX of detecting differentially expressed genes correlated with a cancerous  
XX state of a mammalian cell comprising detecting a gene product encoded by  
XX 65 of the 1079 sequences given in the specification. The polynucleotides  
XX are used to monitor patients having (or susceptible) to cancer to detect  
XX potentially malignant events at a molecular level before they are  
XX detectable at a gross morphological level. The polynucleotides are also

CC useful for monitoring the efficacy of various therapies and preventive  
CC interventions. Polynucleotide probes based on the disclosed sequences  
CC are useful for chromosome mapping and detection of transcription levels.  
CC The 1079 polynucleotide sequences were derived from a human colon cancer  
CC cell line Kml2L4-A cDNA library.

XX Sequence 408 BP; 117 A; 96 C; 98 G; 97 T; 0 other;

Query Match 10.5%; Score 406.4; DB 21; Length 408;

Best Local Similarity 99.8%; Pred. No. 2.5e-72;

Matches 407; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 559 CCGTGTCTCACACCTTTCTACTGAGCATCCTCATGACGAAATGATGAGGAAATGG 618

Db 1 CCGTGTCTCACACCTTTCTACTGAGCATCCTCATGACGAAATGATGAGGAAATGG 60

QY 619 CATATTATAAGAGAGCTGCTGGTGGGAGGACTACATTAAAGACCTGGAAACCAAGTCAT 678

Db 61 CATATTATAAGAGAGCTGCTGGTGGGAGGACTACATTAAAGACCTGGAAACCAAGTCAT 120

QY 679 ATGAAGCCTGTTTCATCCGAGCAGTGGGGCATACAAACGTTGAGAACTGGAGAACTCCA 738

Db 121 ATGAAGCCTGTTTCATCCGAGCAGTGGGGCATACAAACGTTGAGAACTGGAGAACTCCA 180

QY 739 TCACAGACATGGAGCTGGCCCTTCCCGACTTCTTCAAGCCTTTTACGAGTGTCTGCAG 798

Db 181 TCACAGACATGGAGCTGGCCCTTCCCGACTTCTTCAAGCCTTTTACGAGTGTCTGCAG 240

QY 799 CTGCGAGGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCATC 858

Db 241 CTGCGAGGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCATC 300

QY 859 ATTATGTAGAAAGTCTGGAATGCAAAATACAGTGTGAAGAGACCTCACCCAGTTATAG 918

Db 301 ATTATGTAGAAAGTCTGGAATGCAAAATACAGTGTGAAGAGACCTCACCCAGTTATAG 360

QY 919 GAGGTATCCGGTTGAGAAATTTGTGGTACCATTGATCATTTACTTTC 966

Db 361 GAGGTATCCGGTTGAGAAATTTGTGGTACCATTGATCATTTACTTTC 408

## RESULT 14

ABN94955/c

ID ABN94955 standard; DNA; 394 BP.

XX ABN94955;

XX AC

XX 13-AUG-2002 (first entry)

XX Gene #1453 used to diagnose liver cancer.

XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
XX metastatic liver tumor; cytostatic; expression profile; disease state;  
XX disease progression; drug toxicity; drug efficacy; drug metabolism.

XX Homo sapiens.

XX WO200229103-A2.

XX 11-APR-2002.

XX 02-OCT-2001; 2001WO-US30589.

XX 02-OCT-2000; 2000US-237054P.

XX (GENE-) GENE LOGIC INC.

XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX WPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer,

XX hepatocellular carcinoma or metastatic liver tumor in a patient,

PT involves detecting the level of expression of two or more genes in a  
PT liver tissue sample  
XX  
PS Claim 1; SEQ ID NO 1453; 298pp; English.  
XX  
CC The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumour in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN93503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytosolic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pot\_sequences.  
XX  
SQ Sequence 394 BP; 143 A; 65 C; 104 G; 82 T; 0 other;

Query Match 10.1%; Score 392.4; DB 24; Length 394;  
Best Local Similarity 99.7%; Pred. No. 1.6e-69; Mismatches 0; Gaps 0;  
Matches 393; Conservative 0; Indels 0; Gaps 0;  
QY 1709 GATGATCCATCTGCTTGGCTCCACAGTGTGGGATTACAGGCGTGAGCCACCATGCC 1769  
DB 394 GATGATCCATCTGCTTGGCTCCACAGTGTGGGATTACAGGCGTGAGCCACCATGCC 335  
QY 1769 CGGCTCTTCTACACCTTTACACCTGCTCTTATCTCTCTCTCTCTCTCTCTCTCA 1828  
DB 334 CGGCTCTTCTACACCTTTACACCTGCTCTTATCTCTCTCTCTCTCTCTCTCTCA 275  
QY 1829 TCCTGCT 1888  
DB 274 TCCTGCT 215  
QY 1889 TTTTGGTTGAAGGCGAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1948  
DB 214 TTTTGGTTGAAGGCGAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 155  
QY 1949 TTAATTTTAAATAAGAAACAACTTCTAGAGATGATATTTGTGAAACCTCTCTTGGC 2008  
DB 154 TTAATTTTAAATAAGAAACAACTTCTAGAGATGATATTTGTGAAACCTCTCTTGGC 95  
QY 2009 TTAATTTTCCAGATTTTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2068  
DB 94 TTAATTTTCCAGATTTTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 35  
QY 2069 ATAGGCTAATTTCTCCAGCTCAGATGCTCT 2102  
DB 34 ATAGGCTAATTTCTCCAGCTCAGATGCTCT 1

RESULT 15  
AAH50786  
ID AAH50786 standard; cDNA; 401 BP.  
XX  
AC AAH50786;  
XX  
DT 23-AUG-2001 (first entry)  
XX  
XX Human tumour associated cDNA #115.  
XX  
XX Human; cancer specific gene expression; gene therapy;  
KW age related differential expression; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200136685-A2.  
XX

PD 25-MAY-2001.  
XX  
PF 17-NOV-2000; 2000WO-US31809.  
XX  
XX 17-NOV-1999; 99US-0166056.  
PR 17-NOV-1999; 99US-0166106.  
XX  
PA (NYXI-) NYXIS NEURO THERAPIES INC.  
XX  
XX Kroes RA, Moskal JR, Yamamoto H;  
PI WPI; 2001-355647/37.  
DR  
XX Novel nucleic acid molecules differentially expressed in brain cancers,  
XX useful for ascertaining propensity of cell for malignant phenotype or  
XX ascertaining suitability of anti-neoplastic drug candidate -  
XX  
PS Claim 28; Page 53; 82pp; English.

CC The present invention provides the sequences of 184 cDNA fragments which  
CC are differentially expressed in cancer cell depending on the age of the  
CC patient. They can be used to diagnose and identify treatments for  
CC cancers, particularly brain cancers such as haemangioblastoma, teratoma,  
CC haemangioma, glioblastoma, schwannoma, osteoma and pinealoma. The  
CC present sequence is a cancer-associated cDNA of the invention.  
XX  
SQ Sequence 401 BP; 133 A; 72 C; 90 G; 106 T; 0 other;

Query Match 9.6%; Score 371.6; DB 22; Length 401;  
Best Local Similarity 97.7%; Pred. No. 2.5e-65; Mismatches 9; Indels 0; Gaps 0;  
Matches 377; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 2129 AGCACCACCTGAAACCAAGTTGAGTAGAGAGTGTAGAGTGCAGAAATGTGGCTTTTGGC 2188  
DB 14 AGCACCACCTGAAACCAAGTTGAGTAGAGAGTGTAGAGTGCAGAAATGTGGCTTTTGGC 73  
QY 2189 CCACCTTTGCATCTCCAAAATTTACACGGTTGGCCGATCCCATTTAGGACAATGCTTAGT 2248  
DB 74 CCACCTTTGCATCTCCAAAATTTACACGGTTGGCCGATCCCATTTAGGACAATGCTTAGT 133  
QY 2249 TATAAGTCTCCGAGTTGGAAAAGGAAGCCAGAGCTGTCTAGTTTTCATTCTTTCTTT 2308  
DB 134 TATAAGTCTCCGAGTTGGAAAAGGAAGCCAGAGCTGTCTAGTTTTCATTCTTTCTTT 193  
QY 2309 CAGTAAATATTTATTGAGTACTACTGTGTCTAGGCAATTCACCTGGGAACTAGAGATAC 2368  
DB 194 CAGTAAATATTTATTGAGTACTACTGTGTCTAGGCAATTCACCTGGGAACTAGAGATAC 253  
QY 2369 TTACAGAAATTAACAGGGAAGTTCCTGTGTCTATGGAGCTTACATTCTACAGGAGAAA 2428  
DB 254 TTACAGAAATTAACAGGGAAGTTCCTGTGTCTATGGAGCTTACATTCTACAGGAGAAA 313  
QY 2429 GAGATAGCCAATACATAGGAATAAATATATACAAGGTATCATGTAGTGATAATTGCTGTG 2488  
DB 314 GAGATAGCCAATACATAGGAATAAATATATACAAGGTATCATGTAGTGATAATTGCTGTG 373  
QY 2489 GAGAAAAATAAAGCAGGCGAGGAGT 2514  
DB 374 GAAAAAATAAAGCTTTAAGGCGGAAT 399

Search completed: January 30, 2004, 08:32:11  
Job time : 1013.41 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 07:55:54 ; Search time 8124.86 Seconds  
(without alignments)  
11579.600 Million cell updates/sec

Title: US-09-729-674-1  
Perfect score: 3871  
Sequence: 1 ttctctctctctctctctt.....ataaaaaaaaaaaaaaaa 3871

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 1252238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estmu.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_estc.\*
- 9: gb\_est1.\*
- 10: gb\_est2.\*
- 11: gb\_est3.\*
- 12: gb\_est4.\*
- 13: gb\_est5.\*
- 14: gb\_est6.\*
- 15: em\_estfun.\*
- 16: em\_estom.\*
- 17: em\_gss\_hum.\*
- 18: em\_gss\_inv.\*
- 19: em\_gss\_pin.\*
- 20: em\_gss\_vrt.\*
- 21: em\_gss\_fun.\*
- 22: em\_gss\_mam.\*
- 23: em\_gss\_mus.\*
- 24: em\_gss\_pro.\*
- 25: em\_gss\_red.\*
- 26: em\_gss\_png.\*
- 27: em\_gss\_vrl.\*
- 28: gb\_gss1.\*
- 29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1072.2	27.7	1201	9	AL545483
C 2	1012.4	26.2	1200	9	AL571267
C 3	1004.8	26.0	1201	9	AL569764
C 4	985.2	25.5	1196	9	AL517454

5	979.6	25.3	1201	9	AL514488
6	969.8	25.1	1201	9	AL556976
7	964.6	24.9	1201	13	EX425905
C 8	954.8	24.7	1201	13	EX425904
9	954	24.6	1148	13	BU902192
C 10	947.2	24.5	1046	13	EX380662
11	945.8	24.4	1075	9	AL546910
12	944	24.4	1656	11	AK047506
13	934.6	24.2	1201	9	AL541167
C 14	924.6	23.9	1074	9	AL572160
15	924.4	23.9	1652	11	AK017797
16	923	23.8	1201	9	AL517455
C 17	922.8	23.8	1148	9	AL574681
C 18	918.4	23.7	1201	9	AL579089
19	913.4	23.6	1064	12	BM557305
20	896.6	23.2	1053	12	BM471592
21	891	23.0	1201	13	EX437872
C 22	881	22.8	1201	9	AL568770
23	878.4	22.7	1131	12	BM554453
24	873.4	22.6	1201	9	AL554750
25	866.8	22.4	1201	9	AL543090
26	862.4	22.3	1201	9	AL558186
C 27	859.4	22.2	1021	9	AL580676
28	858.8	22.2	1032	12	BM558598
29	857	22.1	939	13	EX386576
C 30	829	21.4	1201	9	AL514487
31	827.2	21.4	1152	9	AL550966
32	824.8	21.3	1040	9	AL558966
33	810.8	20.9	1021	12	BM807166
34	804.6	20.8	1201	13	EX387779
35	801	20.7	846	13	BU838140
36	788	20.4	961	12	BI255097
37	787.6	20.3	944	13	BU856662
38	784.4	20.3	1109	12	BM550012
39	777	20.1	827	12	BI760650
40	773.6	20.0	1083	12	BM920687
41	771.4	19.9	774	9	AL598447
42	770.6	19.9	854	12	BI860855
43	766.8	19.8	899	13	BO437797
44	765.2	19.8	846	12	BI253401
45	759.6	19.6	933	12	BI858585

ALIGNMENTS

RESULT 1  
AL545483  
LOCUS  
DEFINITION AL545483 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0D1G15YK22 5-PRIME, mRNA sequence.  
ACCESSION AL545483  
VERSION AL545483.2 GI:31267318  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
AUTHORS Li W.B., Gruber C., Jessee J., and Polayes D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT On Feb 15, 2001 this sequence version replaced gi:12877964.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 2233.f For more information about this cluster, see  
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D1G15YK2233.f. Contact : Feng Liang Email : fliang@lifetech.com URL :





QY 738 ATACACAGACATGAGCTGGCCCTTCCCGACTTCTTCAAGCCCTTTTACGAGTGTCTCGCA 797  
Db 1193 MTCCTTAAACAAATDGGKSTGCCCTCCCAATGTCATAA - CTTTTACGAKTTTCTCGCA 1135  
QY 798 GCTCGGAGGGTCCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAGAT 857  
Db 1134 SCCK--CGRGGTCCAGGRARCAAGGACTTCAAGGATTTCT-MCTTTCCATAGCAGAY 1078  
QY 858 CATTATGTAGAAGTCTGGGAATCAAAATACAGTGTGAAGAGAACCTCACCCAGTTATA 917  
Db 1077 AATTATGTAGAAGTCTGGGAATCAAAATACAGTGTGAAGAGAACCTCACCCAGTTATA 1021  
QY 918 GGAGGCTATCCGGTCTGAGAAATTTGGCTTACCATGATCATCTTACCTTTCGCTAT 977  
Db 1020 GRA-GCTATCCGGTCTGAGAAATTTGT-GCTTACCATGATCATCTTACCTTTCGCTAT 963  
QY 978 TATAAGTTGAACGACCTGAAGATGACGCCCTTCTGTGAGTGTGATCTGTCTTTTGA 1036  
Db 962 TATAAGTTGAACGACCTGAAGATGACGCCCTTCTGTGAGTGTGATCTGTCTTTTGA 903  
QY 1037 TCAGATGACAGAGTTCATGACGAGAACCTGTGTGTATTTACAGTACCAAGGACACTTG 1096  
Db 902 TCAGATGACAGAGTTCATGACGAGAACCTGTGTGTATTTACAGTACCAAGGACACTTG 843  
QY 1097 GGCCCTCTCGGATGAGCACTTCCAGCCGACCTGAGCCCTTCTGTGAGTGTGATCTGTCTTTTGA 1156  
Db 842 GGCCCTCTCGGATGAGCACTTCCAGCCGACCTGAGCCCTTCTGTGAGTGTGATCTGTCTTTTGA 783  
QY 1157 CACACTCCAGAGGAGTGTATGACTTTCTGTAAGGAAATATAATGATGATGATGAGG 1216  
Db 782 CACACTCCAGAGGAGTGTATGACTTTCTGTAAGGAAATATAATGATGATGATGAGG 723  
QY 1217 AGAAGTTGTGAATATGTGATGACCTTCTGAACTGGAGGAGACAGTACCCACAGC 1276  
Db 722 AGAAGTTGTGAATATGTGATGACCTTCTGAACTGGAGGAGACAGTACCCACAGC 663  
QY 1277 AACCAAGAGACTTCTCTTGGCTTCCAGAAACACAGATTTCTTGTCTCTTTTCCCAACA 1336  
Db 662 AACCAAGAGACTTCTCTTGGCTTCCAGAAACACAGATTTCTTGTCTCTTTTCCCAACA 603  
QY 1337 GCCAGGCTTGTATACCTCAGAGCTTCTTCTTACTTCCAAAGTGAAGGAGGCC 1396  
Db 602 GCCAGGCTTGTATACCTCAGAGCTTCTTCTTACTTCCAAAGTGAAGGAGGCC 543  
QY 1397 CGTCTCTTAAGTGCATGTCAGGGGTGAGCTGCTTCTCTATCTTCAACCTGCA 1456  
Db 542 CGTCTCTTAAGTGCATGTCAGGGGTGAGCTGCTTCTCTATCTTCAACCTGCA 483  
QY 1457 CCTCATGTTCAACCTATCTTCTTACC--TTTTTTTGTGATGAGTCTCGCTCTCTG 1515  
Db 482 CCTCATGTTCAACCTATCTTCTTACC--TTTTTTTGTGATGAGTCTCGCTCTCTG 423  
QY 1516 CCAGGCTGGA-GTGCAATGCGAGTCTCAGCTCACTGCAACCTCGCTCTTGGTTCA 1574  
Db 422 CCAGGCTGGAATGCAATGCGAGTCTCAGCTCACTGCAACCTCGCTCTTGGTTCA 363  
QY 1575 AGCAATCTGCTGCATCAGCTCCCGAGTACCTGGGATACAGGATGTGCCACACGCC 1634  
Db 362 AGCAATCTGCTGCATCAGCTCCCGAGTACCTGGGATACAGGATGTGCCACACGCC 303  
QY 1635 CGGCTAATTTGTATTTTGTAGTAGACCGGGTTTGGCCATGTGGCCAGGCTGTCTCG 1694  
Db 302 CGGCTAATTTGTATTTTGTAGTAGACCGGGTTTGGCCATGTGGCCAGGCTGTCTCG 243  
QY 1695 AACTCTTGACTTCAGATGATCCATCTGCTTGGCTTCCCAAGTGTGGGATTTACAGG 1754  
Db 242 AACTCTTGACTTCAGATGATCCATCTGCTTGGCTTCCCAAGTGTGGGATTTACAGG 183  
QY 1755 TGAGCCACCATGCCCCGGCTCTTTCTCAGCTTTTACCTGTCTTTTATCCTCAATCTG 1814  
Db 182 TGAGCCACCATGCCCCGGCTCTTTCTCAGCTTTTACCTGTCTTTTATCCTCAATCTG 123  
QY 1815 TTTTACACCTTCATCCCTGTCTCTCTC-ATGTTACACTTGTCTTCTCCCATGTTCA 1873

Db 122 TTTTACACCTTCATCCCTGTCTTCTCTATGTTTACACTTGTCTCCCATGTTCTATAG 63  
QY 1874 CTGCTTCTTACCATTTTGGTTTGAAGGAGTCTTCTCTGGCTTGTGTTTTTGTGTTTT 1933  
Db 62 CTGCTTCTTACCATTTTGGTTTGAAGGAGTCTTCTCTGGCTTGTGTTTTTGTGTTTT 3  
QY 1934 CC 1935  
Db 2 CC 1  
RESULT 3  
AL569764/c  
LOCUS  
DEFINITION  
AL569764 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA  
clone CSODI002Y17 3-PRIME, mRNA sequence.  
ACCESSION  
AL569764  
VERSION  
AL569764.2 GI:31291196  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1201)  
Li W.B., Gruber C., Jessee J. and Polayes D.  
Full-length CDNA libraries and normalization  
Unpublished  
COMMENT  
On Feb 16, 2001 this sequence version replaced gi:12925427.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqres@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 2233.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODI002CH09NP1&cluster=2233.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CSODI002CH09NP1.  
Location/Qualifiers  
1. 1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI002Y17"  
/tissue="PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 344 a 242 c 309 g 281 t 25 others  
ORIGIN  
Query Match 26.0%; Score 1004.8; DB 9; Length 1201;  
Best Local Similarity 97.5%; Pred. No. 2.4e-14;  
Matches 1021; Conservative 7; Mismatches 18; Indels 1; Gaps 1;  
QY 922 GCTATCCGGTTGAGAAATTTGTGGCTACCATGATATCTTACTTGCAGTTTGCCTATTATA 981  
Db 1046 GCTATCCAGTGAAGAA-TTGTGGCTACCATGATATCTTACTTGCAGTTTGCCTATTATA 998  
QY 982 AGTTGAACGACCTTGAAGAAATGTCAGCCCTGTGAGTGTGCTTCTTTGATCAGA 1041  
Db 987 AGTTGAACGACCTTGAAGAAATGTCAGCCCTGTGAGTGTGCTTCTTTGATCAGA 928  
QY 1042 ATGCAAGGTTCATGACAGAACTGTGTATTACAGTACCAAGGACACTTGGGGCC 1101  
Db 927 ATGCAAGGTTCATGACAGAACTGTGTATTACAGTACCAAGGACACTTGGGGCC 868  
QY 1102 TCTCGGATGAGCACTTCCAGCCCGACCTGAGAGCTTCACTTCTTTAATGTGACACAC 1161

```
Db 867 TCTCGATGACACTTCAGCCGACCTGAGCAGTTCACTTCTTTAATGTGACACAC 808
QY 1162 TCCAGAGAGAGTGTATGACTTGTCTAGGAAATATAATGATGATGATGAGGAGAAG 1221
Db 807 TCCAGAGAGAGTGTATGACTTGTCTAGGAAATATAATGATGATGATGAGGAGAAG 748
QY 1222 TTGTGGAATATGTGATGACCTCTTGGAACTGGAGGAGACAGCTAGCCACAGCAACCA 1281
Db 747 TTGTGGAATATGTGATGACCTCTTGGAACTGGAGGAGACAGCTAGCCACAGCAACCA 698
QY 1282 AAGAGACTTCTTTGGCGTTTCCAGAAACAGAGATCTTTTCTCTTTTCCACAGCCCA 1341
Db 697 AAGAGACTTCTTTGGCGTTTCCAGAAACAGAGATCTTTTCTCTTTTCCACAGCCCA 628
QY 1342 GGCTGTTGATACCTCAGAGCCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1401
Db 627 GGCTGTTGATACCTAAGAGCCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 568
QY 1402 CTCTAACTGCAATGTCATCAGGGGTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1461
Db 567 CTCTAACTGCAATGTCATCAGGGGTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 508
QY 1462 TGTTCACACCTATCTTTCTACCTTTTCTTGTGATGAGTCTGCTCTCTCTCTCTCTCT 1521
Db 507 TGTTCACACCTATCTTTCTACCTTTTCTTGTGATGAGTCTGCTCTCTCTCTCTCTCT 448
QY 1522 TGGAGTGAATGGACAGTCTCTCAGCTCACTCACTCACTCCGCTCTCTCTCTCTCTCT 1581
Db 447 TGGAGTGAATGGACAGTCTCTCAGCTCACTCACTCACTCCGCTCTCTCTCTCTCTCT 388
QY 1582 CTGCTGCAATGCTCCGAGTACTGCGGATTCAGGATGTCACACAGCCCGGCTAA 1641
Db 387 CTGCTGCAATGCTCCGAGTACTGCGGATTCAGGATGTCACACAGCCCGGCTAA 328
QY 1642 TTTTGTATTTTAGTAGAGAGCGGGTTTGTGCGATGTTGGCAGGCTGCTCTCTCTCTCT 1701
Db 327 TTTTGTATTTTAGTAGAGAGCGGGTTTGTGCGATGTTGGCAGGCTGCTCTCTCTCTCT 268
QY 1702 GACTTCAGATGATCATCTGCTTGGCTTCCGATGTCACAGTGTGGGATTCAGGCGTGA 1761
Db 267 GACTTCAGATGATCATCTGCTTGGCTTCCGATGTCACAGTGTGGGATTCAGGCGTGA 208
QY 1762 CCATGCGCGGCTCTTTCTCACTTTACACCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1821
Db 207 CCATGCGCGGCTCTTTCTCACTTTACACCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 148
QY 1822 ACCTTCATCCCTGCTTCTCTCATGTTCACTTGTCTTCCCATGTTTCATAGTGCCTTT 1881
Db 147 ACCTTCATCCCTGCTTCTCTCATGTTCACTTGTCTTCCCATGTTTCATAGTGCCTTT 88
QY 1882 CTTACCATTTTGGTTGAGGGAGCTCTCTCTGCTGTTTCTCTCTCTCTCTCTCTCTCT 1941
Db 87 CTTACCATTTTGGTTGAGGGAGCTCTCTCTGCTGTTTCTCTCTCTCTCTCTCTCTCT 28
QY 1942 ATCAGTATTTTAAATTAAGAAA 1968
Db 27 ATAAATATWNTWNTAAATAATAAAA 1
```

```
RESULT 4
AL517454/c 1196 bp mRNA linear EST 12-MAY-2003
LOCUS
DEFINITION
AL517454 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
CSODA003YF11 3-PRIME, mRNA sequence.
ACCESSION
AL517454
VERSION
AL517454.2 GI:30534773
KEYWORDS
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1196)
```

AUTHORS  
TITLE  
JOURNAL  
COMMENT

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
On Feb 13, 2001 this sequence version replaced gi:12780947.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 2233.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODA003CC06NP1&cluster=2233.f. Contact :  
Peng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CSODA003CC06NP1.

## FEATURES

## source

1..1196  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODA003YF11"  
/tissue\_type="NEUROBLASTOMA"  
/clone\_lib="Homo sapiens NEUROBLASTOMA"  
/note="Vector: PCWSPORT\_6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the PCWSPORT 6 vector.  
Library was not normalized."  
BASE COUNT 339 a 246 c 312 g 264 t 35 others

## ORIGIN

Query Match 25.5%; Score 985.2; DB 9; Length 1196;  
Best Local Similarity 97.2%; Pred. No. 6.4e-14;  
Matches 1011; Conservative 8; Mismatches 19; Indels 2; Gaps 2;  
QY 885 ATACAGTGTGAGAGAACCTCACCCAGTTATAGGAGCTATCCGGTTGAGAAATTTGTG 944  
Db 1038 AAATACGWTGTGAKAKACCTCACCCAGTTATAGAGCTATCCGGTTGAGAAATTTKT- 980  
QY 945 GTACACATGATATCATTTACTTTGAGTTTGCCTTATTAAGTTGAACGACTGAAGAATGCA 1004  
Db 979 GCTACCATGATATCATTTACTTTGAGTTTGCCTTATTAAGTTGAACGACTGAAGAATGCA 920  
QY 1005 GCCCCTGTGCAGTCAGCTATCTGCTTTTGTATGAGATGACAGGTCTATCAGCAGAAC 1064  
Db 919 GCCCCTGTGCAGTCAGCTATCTGCTTTTGTATGAGATGACAGGTCTATCAGCAGAAC 860  
QY 1065 CTGGTGTATTACAGTATACACAGGAGACACTTGGGCGCTCTCGGATGAGCATTCCAGCCC 1124  
Db 859 CTGGTGTATTACAGTATACACAGGAGACACTTGGGCGCTCTCGGATGAGCATTCCAGCCC 801  
QY 1125 AGACCTGAAGCAGTTTCACTTCTTTAATGTGACCACTCCAGAGGAGCTGTATGACTTT 1184  
Db 800 AGACCTGAAGCAGTTTCACTTCTTTAATGTGACCACTCCAGAGGAGCTGTATGACTTT 741  
QY 1185 GCTAAGGAAATATAATGATGATGATGAGGAGAGTTTGTGAAATATGTGATGACCTC 1244  
Db 740 GCTAAGGAAATATAATGATGATGATGAGGAGAGTTTGTGAAATATGTGATGACCTC 681  
QY 1245 TTGGAATCTGGAGAGACAGCTAGGCCACAGCAACCAAGAGACTTCTCTTGGGTTCA 1304  
Db 680 TTGGAATCTGGAGAGACAGCTAGGCCACAGCAACCAAGAGACTTCTCTTGGGTTCA 621  
QY 1305 GAAACACAGATTTCTTTGCTTTTCCCAACAGCCAGGCTGTTGATACCTCAGAGCCTT 1364  
Db 620 GGAACACAGATTTCTTTGCTTTTCCCAACAGCCAGGCTGTTGATACCTCAGAGCCTT 561  
QY 1365 CTCCTTACTCTCAAAGTGAAGGAGAGCCCGCTCTCTTAACGTGATGTCATCAGGG 1424  
Db 560 CTCCTTACTCTCAAAGTGAAGGAGAGCCCGCTCTCTTAACGTGATGTCATCAGGG 501  
QY 1425 TGAGCCTGCTTCTCTATCTTTCACACCTGCGCAGCTCATGTTTCCACACCTATCTTCTCACC 1484

500	DB	TTTGGCTGCGCTTCCCTATCTTCACACGTGCACCTCAGTCTCACACTATCTTTCTCACC	441
1485	QY	TTTTTTTTGAGATGGAGTCTCGTCTCTTGCCCAGGCTGGAGTGCAATGGCAGTCTTCA	1544
440	DB	TTTTTTTTTGAGATGGAGTCTCGTCTCTTGCCCAGGCTGGAGTGCAATGGCAGTCTTCA	381
1545	QY	GCTCACTGCAACCTCCGCTCTTGCTTCAAGCAATTTCTGTCATCAGCCTCCCGAGTA	1604
380	DB	GCTCACTGCAACCTCCGCTCTTGCTTCAAGCAATTTCTGTCATCAGCCTCCCGAGTA	321
1605	QY	CTTGGATTACAGGCATGTGCCACACGCCCGGCTAATTTTGTAATTTTAGTAGAGACGG	1664
320	DB	CTTGGATTACAGGCATGTGCCACACGCCCGGCTAATTTTGTAATTTTAGTAGAGACGG	261
1665	QY	GGTTTTGCCAATGTGGCCAGGCTGGTCTCGAACTCTTGACATTCAGATGATCCATCTGCCT	1724
260	DB	GGTTTTGCCAATGTGGCCAGGCTGGTCTCGAACTCTTGACATTCAGATGATCCATCTGCCT	201
1725	QY	TGGCTCCCAAGTGTCTGGGATTCAGGCGTAGCGACCATGCCCGGGCTCTTTCTCACC	1784
200	DB	TGGCTCCCAAGTGTCTGGGATTCAGGCGTAGCGACCATGCCCGGGCTCTTTCTCACC	141
1785	QY	TTTACACCTGTCTTCTTATCCTCCATCTGTGTTTTACACCTTCATCCCTGTCTTCCCTCAT	1844
140	DB	TTTACACCTGTCTTCTTATCCTCCATCTGTGTTTTACACCTTCATCCCTGTCTTCCCTCAT	81
1845	QY	GTTACACTGTCTTCCCATGTTACAGTGCCTTTCTTACCAATTTGGTTTCAAGGGC	1904
80	DB	GTTACACTGTCTTCCCATGTTACAGTGCCTTTCTTACCAATTTGGTTTCAAGGGC	21
1905	QY	AGCTTCTCTGGCTGTGTTT	1924
20	DB	AGCTTCTCTGGCTGTGTTT	1

RESULT 5  
AL514488  
LOCUS  
DEFINITION  
AL514488 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone  
CLOB003ZF09 5-PRIME, mRNA sequence.  
VERSION  
AL514488.2 GI:30464373  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
On Feb 13, 2001 this sequence version replaced gi:12777982.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 2233.f For  
more information about this cluster, see  
<http://www.genoscope.cns.fr/cg-bin/cluster.cgi?seq=Clob003ZF09RP1&cluster=2233.f>. Contact :  
Feng liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL :  
<http://fulllength.invitrogen.com/Invitrogen/Corporation1600>  
Faraday Avenue Genoscope sequence ID : CLOB003ZF09RP1.

FEATURES	Location/Qualifiers
source	1..1201
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="CLOB003ZF09"
	/tissue_type="NEUROBLASTOMA"
	/clone_id="Homo sapiens NEUROBLASTOMA"

						/note=-Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."						
BASE COUNT	274 a	299 c	324 g	246 t	58 others							
ORIGIN												
	Query Match	25.3%	Score 979.6;	DB 9;	Length 1201;							
	Best Local Similarity	94.0%;	Pred. No. 8.4e-14;									
	Matches 1025;	Conservative 31;	Mismatches 31;	Indels 4;	Gaps 4;							
Qy	59	CGCGATGAGCCGGGGCGCGGGGGCGCGCGCTAGCGCTGCTGCGTGCGGTGCCTG	118									
Dd	58	CGGGATGAGACCGGGGGCGCGGGGGCGCGCGCTGCTAGCGCTGCTGTCGTTGCCGCTG	117									
Qy	119	CGCGCTGCGCGCGGGCGCGCCCAATACGAACGCTACAGCTTCCTCGTAGCTTCCCACGGGA	178									
Dd	118	CGCGCTGCGCGCNGGGCGGNCACAATACGAACGCTACAGHTTCCGACAGCTTCCCACGGGA	177									
Qy	179	CGAGCTGATCGCGCTCGAGTCGGCTACCAGCACGGCTGACAAAGTACAGCGCGACGA	238									
Dd	178	CGAGCTGATCGCGCTCGAGTCGGCTACCAGCACGGCTGACAAAGTACAGCGCGACGA	237									
Qy	239	CTGCGCCGAGAGGSTRKGGCTACCTGGAGATCAGCCTCGGCTGCACCCGCTTGTGCGCGA	298									
Dd	238	CTGCGCCGAGAGCGTGGGCTACCTGGAGATHAGGCTCGGCTGCACCGCTGTCTGCGCGA	297									
Qy	299	CAGGAGGCGCTTGTCCACCGCAACTCGAGCGCGCGCGCGCGAGCCCCGCGCGCGG	358									
Dd	298	HAGCGAGGCGCTTCTGTACCGCAACTCGAGCGCGCGCGCGCGAGCCCCGCGCGCGG	357									
Qy	359	CCCTCGCAGCTATCCCGAGCTGCGCCCTCTTCGGGGGCGCTGTGCGCGCGCGCATCGCT	418									
Dd	358	CYTCCGAGCTATYCCGAGCTGCGCTCTTCGGGGGCVTGTGCGCGCGCG - ACTGCCT	416									
Qy	419	CAAGCGTGCAGAGCGGCTGCCAGGCTTCCGCGAGTCCAGCGCCAGCGCGAGGTGCT	478									
Dd	417	CAAGCGTGCAGAG - AGGGCTTGCWAGGCHTYYGHAGTWCNAGCHCAGCHCGAGGTGCT	475									
Qy	479	GGCGGACTTCCAGCGCGCGAGCCCTACAAGTTCCCTCGAGTTCGCTTACATTCAAGGCAAA	538									
Dd	476	GGCGGACTTTHAGCGCGCGAGCWCTAAARTCYTGCAGTTCGCTTACATTAGAGAAA	535									
Qy	539	TAAFTCTCCCAAAGCCATCGCGCTGCTCACACCTTTCTACTGAAGCATCTCTGATGACGA	598									
Dd	536	TAAFTATWCCCAARCWATCGCGCTGCTCAAACTTTTACTGAAGAATCTCTGATGAAGA	595									
Qy	599	AATGATGAAGGAGAACATGGCATATTATAGAGCGCTGCTGGTCCGAGAGCTACATTAA	658									
Dd	596	AATGATGAAGGAGAACATGGNAATTATAPARCATGCTGGTCTGAGGACTACATTAA	655									
Qy	659	AGACCTGGAAACCAAGTCATATGAAGCCCTGTTTCATCCGAGCAGTCCGGGCATCAACGG	718									
Dd	656	AGACCTGGAAACCAAGTTATATGAASCWTGTTTCATCCGAGCAGTCCGGGCATCAACGG	715									
Qy	719	TGAGAACTGGAGAACATCCATCAGACATGGAGCTGGCCCTTCCGACTCTTCCAAGC	778									
Dd	716	TGAGAACTGGAGAACATCCATCAGACATGGAGCTGGCCCTTCCGACTCTTCCAAGC	775									
Qy	779	CTTTTACAGTGTCTCCGAGCGCTCCGAGGGTTCAGGGAGATCAAGAGCTTCAAGGATTT	838									
Dd	776	CTTTTACAGTGTCTCCGAGCGCTCCGAGGGTTCAGGGAGATCAAGAGCTTCAAGGATTT	835									
Qy	839	CTACCTTTCCATAGCAGATCANATATGATGAGAGTTCTGGAATGCAAAATAAGTGTGAAGA	898									
Dd	836	CTACCTTTCCATAGCAGATCANATATGATGAGAGTTCTGGAATGCAAAATAAGTGTGAAGA	895									
Qy	899	GAACTCACCCAGTATAGAGGCTATCCGGTTGAGAAATTTCTGGCTACCATGTATCA	958									
Dd	896	GAACTCACCCAGTATAGAGGCTATCCGGTTGAGAAATTTCTGGCTACCATGTATCA	955									
Qy	959	TTACTTGCAGTTTGCCTATTATTAAGTTGAAACGACCTCAAGAAATCGAGCCCTGTGTCAGT	1018									

956 TTACTTGCAGTTTGGCTTATTATTAAGTTGAACGACCTTGAAGATTCAG-CCCTGTGSCAGT 1014  
 1019 CAGCTATCTGCTTTTGTATCAGATGACAAAGTTCATGACAGCAAGCTTGGTGTATTACCA 1078  
 1015 CAGCTATCTGCTTTTGTATCAGATGACAAAGTTCATGACAGCAAGCTTGGTGTATTACCA 1074  
 1079 GTACACAGGACACATTGGGGCTCTTCGATGAGCACTTCCAGCCGACAGCTGAAGCAGT 1138  
 1075 GTWCCACAGGACACATTGGGGCTTCTCGATGAGCMTT-CMGCCACARACTTGAAGCAGT 1133  
 1139 TCAGTTCTTTA 1149  
 1134 YCAGTTTAA 1144  
  
 AL556976 1201 bp mRNA linear EST 31-MAY-2003  
 AL556976 Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens cDNA  
 clone CS0DH005Y101 5-PRIME, mRNA sequence.  
 AL556976  
 AL556976.2 GI:31278776  
 Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1. (bases 1 to 1201)  
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished  
 On Feb 15, 2001 this sequence version replaced gi:12900137.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 was not normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 2233.f For more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DH005AE01QP1&cluster=2233.f. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DH005AE01QP1.  
 FEATURES  
 source  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DH005Y101"  
 /tissue\_type="T CELLS (JURKAT CELL LINE)"  
 /cell\_line="JURKAT CELL LINE"  
 /clone\_lib="Homo sapiens T CELLS (JURKAT CELL LINE)"  
 /note="vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."  
 BASE COUNT 272 a 339 c 320 g 234 t 36 others  
 ORIGIN  
 Query Match 25.1%; Score 969.8; DB 9; Length 1201;  
 Best Local Similarity 97.7%; Pred. No. 1.3e-13;  
 Matches 1015; Conservative 9; Mismatches 11; Indels 4; Gaps 4;  
 65 GGAGCCGGGGCGCGGGGGCGCGGGCGCTGCTAGCGCTGCTGTGCGTGGCTTGGCGGCT 124  
 58 GGATCCGGGGCGCGGGGGCGCGGGCGCTGCTAGCGCTGCTGTGCGTGGCTTGGCGGCT 117  
 125 GCGGGCGGGGGCGCGGGCGCGCGCGCTTCCGAGCTTCCAGCGGACGAGCT 184  
 118 GC-GCGGGGGCGG-CCAATACGAACGCTACAGCTTCCGAGCTTCCAGCGGACGAGCT 175

185 GATGCGGCTCAGTTCGGCTTACCGGCAACGCGCTTGGACAAAGTACAGCGGAGACCTGGGC 244  
 176 GATGCGGCTCAGTTCGGCTTACCGGCAACGCGCTTGGACAAAGTACAGCGGAGACCTGGGC 235  
 245 CGAGAGCGTGGCTTACCTGGAGATCAGCGCTTGGCGTGCACCGCTTGTCTGCGCGACAGCGA 304  
 236 CGAGAGCGTGGCTTACCTGGAGATCAGCGCTTGGCGTGCACCGCTTGTCTGCGCGACAGCGA-A 234  
 305 GGCCTTCTGCGACCGCAACTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 364  
 295 GGCCTTCTGCGACCGCAACTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 354  
 365 CAGCTATCCGAGCTGCGCTTCTGGGGGCGCTTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 424  
 355 CAGCTATCCGAGCTGCGCTTCTGGGGGCGCTTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 414  
 425 CTGCAAGCAGCGGCGCTTCCGCGAGTCCGCGAGTCCGCGAGCGCGCGCGCGCGCGCGCGCG 484  
 415 CTGCAAGCAGCGGCGCTTCCGCGAGTCCGCGAGTCCGCGAGCGCGCGCGCGCGCGCGCGCG 474  
 485 CTTCCAGCGCGCGAGCGCTTCAAGTTCTTCCGAGTTCGCTTCTTCAAGCAAAATATCT 544  
 475 CTTCCAGCGCGCGAGCGCTTCAAGTTCTTCCGAGTTCGCTTCTTCAAGCAAAATATCT 534  
 545 CCCCAAGCGCATCGCGCTTCTCACACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 604  
 535 CCCCAAGCGCATCGCGCTTCTCACACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 594  
 605 GAAGAGAAATGCGCATATTATTAAGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 664  
 595 GAAGAGAAATGCGCATATTATTAAGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 654  
 665 GGAACCAAGTTCATTAAGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 724  
 655 GGAACCAAGTTCATTAAGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 714  
 725 CTGAGAGCAATCCATCAAGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 784  
 715 CTGAGAGCAATCCATCAAGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 774  
 785 CGAGTGTCTGCGAGCGCTTCCAGGGAGTTCAGGGAGTTCAGGGAGTTCAGGGAGTTCAGGGAG 844  
 775 CGAGTGTCTGCGAGCGCTTCCAGGGAGTTCAGGGAGTTCAGGGAGTTCAGGGAGTTCAGGGAG 834  
 845 TTCCATAGCAGATCATTTATGTAGAGTTCTTGAATGCAAAATACAGTGTGAAGAGACCT 904  
 835 TTCCATAGCAGATCATTTATGTAGAGTTCTTGAATGCAAAATACAGTGTGAAGAGACCT 894  
 905 CACCCAGTTATAGAGCTTCCGGTTGAGAAATTTTGGCTACCATGTATCATTTACTT 964  
 895 CACCCAGTTATAGAGCTTCCGGTTGAGAAATTTTGGCTACCATGTATCATTTACTT 954  
 965 GCAGTTTCTCTTATTAAGTTGAACGACCTGAAGATGCGAGCGCGCGCGCGCGCGCGCGCGCTA 1024  
 955 GCAGTTTCTCTTATTAAGTTGAACGACCTGAAGATGCGAGCGCGCGCGCGCGCGCGCGCGCTA 1014  
 1025 TCTGCTTTTCTGATCAGAAATGCAAGGCTCATGCGAGCAAACTGTGTATTACAGTACCA 1084  
 1015 TCTGCTTTTCTGATCAGAAATGCAAGGCTCATGCGAGCAAACTGTGTATTACAGTACCA 1073  
 1085 CAGGGACACTTGGGGCGCTC 1103  
 1074 MAGGACACTTGGGGCGCTC 1092

RESULT 7  
 BX425905  
 LOCUS  
 DEFINITION  
 CS0DA004Y407 5-PRIME, mRNA sequence.  
 ACCESSION  
 BX425905  
 VERSION  
 BX425905.1 GI:30774508  
  
 BX425905 1201 bp mRNA linear EST 15-MAY-2003  
 BX425905 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone  
 CS0DA004Y407 5-PRIME, mRNA sequence.





FEATURES	Location/Qualifiers	source	
	1. .1201		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS0DA004YM07"		
	/tissue_type="NEUROBLASTOMA"		
	/clone_lib="Homo sapiens NEUROBLASTOMA"		
	/note="Vector: pCMVSPORT6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."		
BASE COUNT	309 a 234 c 320 g 275 t	63 others	
ORIGIN			
	Query Match	24.7%; Score 954.8; DB 13; Length 1201;	
	Best Local Similarity	90.6%; Pred. NO. 2.8e-13;	
	Matches 1006; Conservative	27; Mismatches 74; Indels 3; Gaps 2;	
QY	824 GGACTTCAGAGTTTCTACCTTTCACATACAGATCATATTAGAGTTTCGGAAATGCA 883		
DB	1110 GGGGTCGAGGAYAGRCCTTAAAGTTTCCCTTCAARSAAYCATWIKARRAATYKG 1051		
QY	884 AATPACAGTGTGAAGAACCTCACCCAGATTATAGAGGCTATCCGGTTGAGAAATTTGT 943		
DB	1050 AATGCAAAWACAPKGAARAAACCCCCCAKTAWRGAGCTACCCGGTGAGAAATTKG 991		
QY	944 GGTACCATGTATCATTAAGTTCAGTTTGGCTTATTAAGTTGAACGACCTGAAGATGC 1003		
DB	990 GGTACCATGTATCATTAAGTTCAGTTTGGCTTATTAAGTTCGAACGACCTGAAGATGC 934		
QY	1004 AGCCCTCTGTCAGTATCTGCTTTTATGATCAGATGACAAAGGTCATGACAGCAA 1053		
DB	933 AGCCCTCTGTCAGTATCTGCTTTTATGATCAGATGACAAAGGTCATGACAGCAA 874		
QY	1064 CTTGGTGTATTACAGTACACAGGAGACCTTGGGCGCTCTCGATGAGCATTCCAGCC 1123		
DB	873 CTTGGTGTATTACAGTACACAGGAGACCTTGGGCGCTCTCGATGAGCATTCCAGCC 814		
QY	1124 CAGACCTGAAGCAGTTTCAAGTTTAAATGTGACACACTCCAGAGGAGCTGTATGACTT 1183		
DB	813 CAGACCTGAAGCAGTTTCAAGTTTAAATGTGACACACTCCAGAGGAGCTGTATGACTT 754		
QY	1184 TGTAAAGGAAATATTAATGGATGATGATGAGGAGAAAGTTGGAAATATGTGGATGACT 1243		
DB	753 TGTAAAGGAAATATTAATGGATGATGATGAGGAGAAAGTTGGAAATATGTGGATGACT 694		
QY	1244 CTTGGACTGGAGAGACAGCTAGCCACAGCAACCAAGAGACTTCTCTTTGGCGTTC 1303		
DB	693 CTTGGAACTGGAGAGACAGCTAGCCACAGCAACCAAGAGACTTCTCTTTGGCGTTC 634		
QY	1304 AGGAAACACAGATTCTTTTGTCTTTTCCCAACAGCCAGGCTTGTGATACCTCAGAGCT 1363		
DB	633 AGGAGACACAGATTCTTTTGTCTTTTCCCAACAGCCAGGCTTGTGATACCTCAGAGCT 574		
QY	1364 TCTCTTTACTCTCCAAAGTGAAGGAGAGCCCGCTCTCTAACTGCAATGTCATCAGG 1423		
DB	573 TCTCTTTACTCTCCAAAGTGAAGGAGAGCCCGCTCTCTAACTGCAATGTCATCAGG 514		
QY	1424 GTGAGCTGCTCTTCTTATCTTACACCTTCCACCTCATGTTTACACCTTATCTTCTCAC 1483		
DB	513 GTGAGCTGCTCTTCTTATCTTACACCTTCCACCTCATGTTTACACCTTATCTTCTCAC 454		
QY	1484 CTTTTTTTTGAGATGAGTCTCGCTCTCTTGGCCAGGCTGGAGTGAATGGACGTTCTC 1543		
DB	453 CTTTTTTTTGAGATGAGTCTCGCTCTCTTGGCCAGGCTGGAGTGAATGGACGTTCTC 394		
QY	1544 AGCTCACTGCAACTCCGCTCTTTGGGTTTCAAGCAATTTCTGCTGATCAGCTCCCGAGT 1603		
DB	393 AGCTCACTGCAACTCCGCTCTTTGGGTTTCAAGCAATTTCTGCTGATCAGCTCCCGAGT 334		
QY	1604 ACCTGGGATPACAGGCAATGTCACACACGCCCGGCTAAATTTTGTATTTTATAGACG 1663		



```
QY 127 GCGCGGCGCGCGCCCAATACGAACCTACAGTTCCTCCGAGCTTCCCGGACGAGCTGA 186
DB 61 GCGCGGCGCGCGCCCAATACGAACCTACAGTTCCTCCGAGCTTCCCGGACGAGCTGA 120
QY 187 TCGCGCTGAGTCGCGCTACCGGACGCGCTGGACAGTACAGCGGCGAGCACTGGGCGG 246
DB 121 TCGCGCTGAGTCGCGCTACCGGACGCGCTGGACAGTACAGCGGCGAGCACTGGGCGG 180
QY 247 AGAGCTKGGCTACCTGGAGATCAGCTGCGGCTGCACCGTTCCTGGCGACAGCGAGG 306
DB 181 AGAGCTKGGCTACCTGGAGATCAGCTGCGGCTGCACCGTTCCTGGCGACAGCGAGG 240
QY 307 CTTCTGCGACGCGAATCGAGCGCGCGCGAGCGCGCGAGCGCGCGCGCGCGCTCGCCA 366
DB 241 CTTCTGCGACGCGAATCGAGCGCGCGCGAGCGCGCGAGCGCGCGCGCGCGCTCGCCA 300
QY 367 GCTATCCGAGCTGCGCTCTTCGGGGGCGCTGCTCGCGCGCGCGAGCTGCTCAAGGCT 426
DB 301 GCTATCCGAGCTGCGCTCTTCGGGGGCGCTGCTCGCGCGCGAGCTGCTCAAGGCT 360
QY 427 GCAAGCAGGCGCTGCGAGCTTCGCGCAGTCCAGCGCGCGAGCGCGAGGTGTGCGGACT 486
DB 361 GCAAGCAGGCGCTGCGAGCTTCGCGCAGTCCAGCGCGCGAGCGCGAGGTGTGCGGACT 420
QY 487 TCCAGCGCGCGAGCGCTTCAAGTTCCTGCGAGTTCGCTTACTTCAAGCGCAATATCTCC 546
DB 421 TCCAGCGCGCGAGCGCTTCAAGTTCCTGCGAGTTCGCTTACTTCAAGCGCAATATCTCC 480
QY 547 CAAAGCAGCTGCGCGCTGCTCACCTTCTACTGAAGCATCTCTGATGACGAATGATGA 606
DB 481 CCAAGCAGCTGCGCGCTGCTCACCTTCTACTGAAGCATCTCTGATGACGAATGATGA 540
QY 607 AGAGAACATGCGATATATAGAGCTGCTGCTGCGCGAGCTACATTAAGACCTGG 666
DB 541 AGAGAACATGCGATATATAGAGCTGCTGCTGCGCGAGCTACATTAAGACCTGG 600
QY 667 AAACCAAGTCATATGAAGCTGTTTCATCCGAGCAGTGGCGGATACACGCTGAGAACT 726
DB 601 AAACCAAGTCATATGAAGCTGTTTCATCCGAGCAGTGGCGGATACACGCTGAGAACT 660
QY 727 GGAGAACATTCATCAGACATGAGCTGGCGCTTCCGAGCTTCTTCAAGCTTTTACG 786
DB 661 GGAGAACATTCATCAGACATGAGCTGGCGCTTCCGAGCTTCTTCAAGCTTTTACG 720
QY 787 AGTGTCTCGAGCTGCGAGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTACCTTT 846
DB 721 AGTGTCTCGAGCTGCGAGGTTCCAGGAGATCAAGGACTTCAAGGATTTCTACCTTT 780
QY 847 CCATAGCAGATCATTTATGTAAGTTCCTGGAATGCAAAATACAGTGTGAAGAGAACCTCA 906
DB 781 CCATAGCAGATCATTTATGTAAGTTCCTGGAATGCAAAATACAGTGTGAAGAGAACCTCA 840
QY 907 CCCAGTTATAGAGGCTATCCGTTGAGAAATTTGCGCTACCATGATCATTTACTTGC 966
DB 841 CCCAGTTATAGAGGCTATCCGTTGAGAAATTTGCGCTACCATGATCATTTACTTGC 900
QY 967 AGTTTGCCTATTATAGTTGAAGCAGCTGAAGAAATGCG - CCCCCTGTGCGAGTCACTAT 1025
DB 901 AGTTTGCCTATTATAGTTGAAGCAGCTGAAGAAATGCGCCGCCCTCTGTGAGGCCAGTAT 960
QY 1026 CTGCTCTTTGATCAGAAATGCAAGGTCATGCAGC - AGAAGCTGGTGTATTACAGTACG 1083
DB 961 CGGCTCTTGTATCAGAAAGACAGGCTCTGCGCCCAACAACTGGGGGTATTACCGGCTCC 1020
QY 1084 ACAGGGACAC - - - TTGGGGCTCTCGGA 1108
DB 1021 CCGCGGAGACCTTTTGGGGCCCTTCGGA 1048
```

RESULT 10  
BX380662/c  
LOCUS BX380662 1046 bp mRNA linear EST 08-MAY-2003

DEFINITION  
BX380662 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0D1056Y117 3-PRIME, mRNA sequence.  
BX380662  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1046)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
COMMENT  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 2233.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgt-bin/cluster.cgi?seq=CS0D1056AE09NP1&cluster=2233.f. Contact :  
Feng Liang Email : fliang@life.techn.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0D1056AE09NP1.

FEATURES  
Location/Qualifiers  
1..1046  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1056Y117"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 304 a 210 c 286 g 226 t 20 others  
ORIGIN

Query Match 24.5% Score 947.2; DB 13; Length 1046;  
Best Local Similarity 97.9%; Pred No 4.5e-13;  
Matches 953; Conservative 10; Mismatches 9; Indels 1; Gaps 1;

```
QY 969 TTTGCCCTATTATAGTTGAACGACCTGAGAAATCGAGCCCTGTGCAGTCAGCTATCTG 1028
DB 980 KTTGCCCTTTTATAGTTGAACGACCTGAGAAAGCCTGTCADTGCAGCTWTCTG 921
QY 1029 CTCTTTGATCAGATGACAGAGTCTGACAGCAGACCTGCTGATTACCAAGTACCAGG 1088
DB 920 CTCTTTGATCAGATGACAGAGTCTGACAGCAGACCTGCTGATTACCAAGTACCAGG 861
QY 1089 GACACTTGGGCTCTCGGATGACACTTCCAGCCAGACCTGGAAGCAGTTCAGTTCTTT 1148
DB 860 RACACTTGGGCTCTCGGATGACACTTCCAGCCAGACCTGGAAGCAGTTCAGTTCTTT 801
QY 1149 AATGTGACCACTCCAGAGAGGCTGTATGACTTTGCTAAGGAAATATATGATGAT 1208
DB 800 AATGTGACCACTCCAGAGAGGCTGTATGACTTTGCTAAGGAAATATATGATGAT 741
QY 1209 GATGAGGAGAGTGTGGAAATATGATGATGACCTCTTGAAGCTGGAGGAGACAGCTAG 1268
DB 740 GATGAGGAGAGTGTGGAAATATGATGATGACCTCTTGAAGCTGGAGGAGACAGCTAG 681
QY 1269 CCCACAGCAACCAAGAGACTTCTCTTGGCGTTCAGGAAACACAGATTTCTTTGTCCTTT 1328
DB 680 CCCACAGCAACCAAGAGACTTCTCTTGGCGTTCAGGAAACACAGATTTCTTTGTCCTTT 621
QY 1329 TCCCAAGCGCCGAGCTGTGATACCTCAGAGCTCTCTTTACTCTCCAAAGTGAAGG 1388
DB 620 TCCCAAGCGCCGAGCTGTGATACCTCAGAGCTCTCTTTACTCTCCAAAGTGAAGG 561
QY 1389 GAAGCCCGCTCTCTCTAACTGCTATCATCAGGGGTGAGCCTGCTTCTCTCTTCTAC 1448
```





```
Db 9 TTCCTAGGCTGCGCGCTCGCGGTCTCTAGCTCGGGCTCTGTGTTCTGTGGCCACGCGC 68
Qy 62 GATGAGCCCGGGGCGCGGGGCGCGGCTGCTAGCGCTGCTGTGGGTGGCTGCGC 121
Db 69 GATGGGCCCCGAGCCCTACGCCC---GGTTGCTGGTCTGCTGTGTGGTGGC 125
Qy 122 GCTGCGCGCGGGGCGCGCCCAATACGAACCTACAGCTTCGGAGCTTCCACGGACGA 181
Db 126 GCGACACCCCGGGCGGGGAGTAGACGAGCTACAGCTTCGGAGCTTCCGCGGACGA 185
Qy 182 GCTGATCCGCTCGAGTCGCGCTACCGGACGCGCTGGACAGTACAGCGCGGAGCACTG 241
Db 186 GCTGATCCGCTCGAGTCGCGCTACCGGACGCGCTGCACAGTACAGCGCGGAGCACTG 245
Qy 242 GCGCGAGAGSTKGGCTACTGAGATCAGCTTCGGGCTGCGCGCTGCTGCGCGACAG 301
Db 246 GCGCGAGAGCTGGGCTACCTGGAGTGAAGCTGCGGCTGCGCGCTGCTGCGCGACAG 305
Qy 302 CGAGGCGCTTTCGCCACGCAACTGACGCGCGCGCGCGAGCGCGCGCGCGCGCT 361
Db 306 CGAGGCGCTTTCGCCACGCAACTGACGCGCGCGCGCGAGCGCGCGCGCGCGCGC 365
Qy 362 CGCGAGTATCCCGAGCTGCGCTCTTTCGGGCGGCTGCTGCGCGCGCGCGCTGCTCAA 421
Db 366 GCGCGAGCGCGCGAGCTTTCGGGCGGCTGCTGCGCGCGCGCGCGCTGCTCAA 425
Qy 422 GCGCTGCAAGCAGGCGCTGCGCGCTTTCGGGCGGCTGCTGCGCGCGCGCGCTGCTG 481
Db 426 GCGCTGCAAGCAGGCGCTGCGCGCTTTCGGGCGGCTGCTGCGCGCGCGCGCTGCTG 485
Qy 482 GGACTTCCAGCGCGCGAGCGCTTCAAGTTCCTGCGAGTTCGCTTCAAGCAATAA 541
Db 486 GGACTTCCAGCGCGCGAGCGCTTCAAGTTCCTGCGAGTTCGCTTCAAGCAATAA 545
Qy 542 TCTCCCAAGCCATGCGCGTGTCTACACTTCTTCTAGAGCTGCTGAGCGCGCGAGT 601
Db 546 CTTCCCAAGCCATGCGCGTGTCTACACTTCTTCTAGAGCTGCTGAGCGCGCGAGT 605
Qy 602 GATGAAGAGGAACATGCGATATTAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 661
Db 606 GATGAAGAGGAACATGCGATATTAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 665
Qy 662 CTTGGAACCAAGTCAATGAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 721
Db 666 CTTGGAACCAAGTCAATGAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 725
Qy 722 GACTGGAGACATCCATCAGACATGAGCTGCGCTTCCGAGCTTCTTCAAGCCTT 781
Db 726 GAACTGGAGAACGTCCTATTCGACATGAGCTGCGCTTCCGAGCTTCTTCAAGCCTT 785
Qy 782 TTACGAGTGTCTCGACGCTGCGAGGCTTCCAGGAGATCAAGGACTTCAAGGATTCTA 841
Db 786 CTACGAGTGTCTGCTGCTGCTGCGAGGCTGCGGAGATCAAGGACTTCAAGGACTTCTA 845
Qy 842 CTTTCCATAGACATCATTTATGAGTCTGGAATGCAAAATACAGTGTGAAGAGAA 901
Db 846 CTTTCCATAGACATCATTTATGAGTCTGGAATGCAAAATACAGTGTGAAGAGAA 905
Qy 902 CTTCAACCCAGCTTATAGAGGCTATCCGCTTGAGAAATTTGCTGCTACCATGATCATTA 961
Db 906 CTTCAACCCAGCTATAGAGGCTATCCGCTTGAGAAATTTGCTGCTACCATGATCATTA 965
Qy 962 CTTGAGTGTCTTATTAAGTGAAGCTGGAAGATGCAAGCTTCTGCTGCTGCTGCTGCTG 1021
Db 966 TTTACAGTGTGCTTATTAAGTGAAGCTGGAAGATGCAAGCTTCTGCTGCTGCTGCTGCTG 1025
Qy 1022 CTATCTGCTTCTTATCAGAAATCAAGGCTATCAGCAGAACTGCTGCTGCTGCTGCTGCTG 1081
Db 1026 CTACTGCTTCTTGAACAGAGTGAAGGCTATCAGCAGAACTGCTGCTGCTGCTGCTGCTG 1085
Qy 1082 CCACAGGAGACTTTGGGCGCTTCTCGGATGAGCACTTCCAGCCCGAGCTTGAAGAGCTTCA 1141
Db 1086 .CCACGGGAGCAAGTGGGCGCTTCTCGGATGAGCACTTCCAGCCCGAGCTTGAAGAGCTTCA 1145
```

```
Qy 1142 GTTCTTTAATGTACCACTCCAGAGGAGCTGTATGACTTTGCTAAGGAAATATAAT 1201
Db 1146 GTTCTTTAATGTACCACTCCAGAGGAGCTGTATGACTTTGCTAAGGAAATATAAT 1205
Qy 1202 GGATGATGATGAGGAGAGTGTGGAATATGTGATGAGCTCTTGGAACTTGGAGGAGAC 1261
Db 1206 GGATGATGATGAGGAGAGTGTGGAATATGTGATGAGCTCTTGGAACTTGGAGGAGATC 1265
Qy 1262 CAGCTAGCCACAGCAACCAAGAGACTTCT 1293
Db 1266 TGCCTAGTCCACAGGGCTAAGGAACCTCTCT 1297
```

## RESULT 13

AL541167

LOCUS

DEFINITION

AL541167 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE005YF12

ACCESSION

AL541167

VERSION

AL541167.2 GI:30545079

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 1201)

AUTHORS

Li, W.B., Gruber, C., Jesses, J., and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished

COMMENT

On Feb 15, 2001 this sequence version replaced gi:12871971.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 2233.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DE005DC06QPI&amp;cluster=2233.f. Contact :

Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/

Faraday Avenue Genoscope sequence ID: CS0DE005DC06QPI.

Location/Qualifiers

1..1201

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DE005YF12"

/tissue\_type="PLACENTA"

/clone\_lib="Homo sapiens PLACENTA"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

BASE COUNT 258 a 335 c 313 g 237 t 58 others

ORIGIN

Query Match 24.2%; Score 935.6; DB 9; Length 1201;

Best Local Similarity 98.2%; Pred. No. 7.2e-13;

Matches 976; Conservative 10; Mismatches 4; Indels 4; Gaps 4;

Qy 57 GCGCGATGAGAGCG 116

Db 58 GCGCGATGAGAGCG 117

Qy 117 TCGCGCTGCG 176

Db 118 TCGCGCTGCG 176

Qy 177 GACGAGCTGATCGCGCTGAGTTCGCGCTACCGGACGCGCTGCGGACGCGCTGCGG 236

Db 178 GACGAGCTGATCGCGCTGAGTTCGCGCTACCGGACGCGCTGCGGACGCGCTGCGG 236

```
Db 177 GAGAGCTGATCGCGCTGAGTGGCTTACCGGACGCGCTGGACAAAGTACAGCGCGAG 236
QY 237 CACTGGCGGAGAGCTGGCTACCTGGAGATCAGCTCGCGCTGCACCGCTTGTGCGC 296
Db 237 CACTGGCGGAGAGCTGGCTTACCTGGAGATCAGCTCGCGCTGCACCGCTTGTGCGC 296
QY 297 GACAGGAGGCTTCTGCGCACCGCACTGACGCGCGCGCGCGCGCGCGCGCGCGC 356
Db 297 GACAGC-AGGCTTCTGCGCACCGCACTGACGCGCGCGCGCGCGCGCGCGCGCGC 355
QY 357 GGCCTCGCAGCTATCCCGAGCTGCGCTCTTCTGCGGCGCTCTGCGCGCGCGCACTGC 416
Db 356 GGCCTCGCAGCTATCCCGAGCTGCGCTCTTCTGCGGCGCTCTGCGCGCGCGCACTGC 415
QY 417 CTCAGCGCTGCAAGCAGGCGCTGCGCGCTTCTGCGGCGCTGCGCGCGCGCGCGAGTG 476
Db 416 CTCAGCGCTGCAAGCAGGCGCTGCGCGCTTCTGCGGCGCTGCGCGCGCGCGAGTG 475
QY 477 CTGGCGGAGCTTCCAGCGCGCGCGCTTACAGCTTCTGCGCTTCTGCGCTTCTGCGCA 536
Db 476 CTGGCGGAGCTTCCAGCGCGCGCGCTTACAGCTTCTGCGCTTCTGCGCTTCTGCGCA 535
QY 537 AATAATCTCCCAAGCCATGCGCGCTGCTCAACCTTCTACTGAAGCATCTGATGAC 596
Db 536 AATAATCTCCCAAGCCATGCGCGCTGCTCAACCTTCTACTGAAGCATCTGATGAT 595
QY 597 GAAATGATGAAGAGGAACATGATTAAGAGCGCTGCGCTGCTGCGGAGGACTACATT 656
Db 596 GAAATGATGAAGAGGAACATGATTAAGAGCGCTGCGCTGCTGCGGAGGACTACATT 655
QY 657 AAAGACCTGGAACCAAGTCATATGAAGCCTGTTTATCCGAGCAGTGGCGGATACAA 716
Db 656 AAAGACCTGGAACCAAGTCATATGAAGCCTGTTTATCCGAGCAGTGGCGGATACAA 715
QY 717 GGTGAGACTGGAGACATCCATCAGACATGAGCTGCGCTTCCCGACTTCTTCAA 776
Db 716 GGTGAGACTGGAGACATCCATCAGACATGAGCTGCGCTTCCCGACTTCTTCAA 775
QY 777 GCCTTTTACAGTGTCTCGAGCGCTCGAGGGTTCCAGGGAGATCAAGGACTTCAAGGAT 836
Db 776 GCCTTTTACAGTGTCTCGAGCGCTCGAGGGTTCCAGGGAGATCAAGGACTTCAAGGAT 835
QY 837 TTCTACCTTCCATAGACATATATGATGAGAGTCTGGAATGCAAAATAGAGTGTGA 896
Db 836 TTCTACCTTCCATAGACATATATGATGAGAGTCTGGAATGCAAAATAGAGTGTGA 895
QY 897 GAGAACTACCCAGCTATAGAGGCTATCCGGTTGAGAAATTTGTGCTACCATGAT 956
Db 896 GAGAACTACCCAGCTATAGAGGCTATCCGGTTGAGAAATTTGTGCTACCATGAT 955
QY 957 CATTACTTGCAGTTTGCCTATTATAGTTGAAGCAGCTGAAGAAATGCAAGCGCTGTGCA 1016
Db 956 CATTACTTGCAGTTTGCCTATTATAGTTGAAGCAGCTGAAGAAATGCAAGCGCTGTGCA 1013
QY 1017 GTGAGTATCTGCTTCTGATCAGATGACAGG 1050
Db 1014 GTGAGTATCTGCTTCTGATGAGATGADAGG 1047

RESULT 14
AL572160/c 1074 bp mRNA linear EST 31-MAY-2003
LOCUS AL572160 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA.
DEFINITION clone CSODI032YH24 3-PRIME, mRNA sequence.
ACCESSION AL572160
VERSION AL572160.2 GI:31293545
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1074)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
```

TITLE  
JOURNAL  
COMMENT

Full-length cDNA libraries and normalization

Unpublished

On Feb 16, 2001 this sequence version replaced gi:12930160.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 2233.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSODI032DD12NP1&cluster=2233.f. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Paradise Avenue Genoscope, sequence ID : CSODI032DD12NP1.

Location/Qualifiers

1. 1074

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CSODI032YH24"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/notes="1st strand cDNA was primed with a NotI-cilgo(dt)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 325 a 204 c 283 g 237 t 25 others

ORIGIN

Query Match 23.9%; Score 924.6; DB 9; Length 1074;

Best Local Similarity 97.0%; Pred. No. 1.3e-12;

Matches 971; Conservative 10; Mismatches 16; Indels 4; Gaps 4;

QY 1307 AAACACAGATCTTTGTCCTTTCCACAGCCGAGCTTGTATACCTCAGAGCCTTCT 1366

Db 999 AGAACACAGATCTTTTCCCTTTCCCAACAGCCGAGCT-KTGATACCTCAGAGCCTTCT 941

QY 1367 CTTTACTCTCAAGTGAAGGAGGAGCCCGCTCTCTTAACCTGCATGCATCAGGGGTG 1426

Db 940 CTTTACTCTCAAGTGAAG- GGAAGCCCGCTCTCTTAACCTGCATGCATCAGGGGTG 892

QY 1427 AGCTGCTTTCCTATCTTACACCTGCGACCTCATGTTTACACCTATCTTTCTACCTT 1486

Db 881 AGCTGCTTTCCTATCTTACACCTGCGACCTCATGTTTACACCTATCTTTCTACCT-T 823

QY 1487 TTTTGTGAGTGGAGTCTCGCTCTCTGCGAGCTGGAGTGAATGGCAGCTTCTCAGC 1546

Db 822 TTTTGTGAGTGGAGTCTCGCTCTCTGCGAGCTGGAGTGAATGGCAGCTTCTCAGC 763

QY 1547 TCACCTGAACCTCCGCTCTTGGGTTCAGCAATTCGTGCAATCAGCTTCCGAGTACC 1606

Db 762 TCACCTGAACCTCCGCTCTTGGGTTCAGCAATTCGTGCAATCAGCTTCCGAGTACC 703

QY 1607 TGGGATTACAGGATGCGCACCGCGCTAAATTTCTATTTTAGTAGAGAGCGGG 1666

Db 702 TGGGATTACAGGATGCGCACCGCGCTAAATTTCTATTTTAGTAGAGAGCGGG 643

QY 1667 TTTTGCCATGTTGCGCAGGCTGCTCGAACTCTTGACTTCAGATGATCCTGCTTG 1726

Db 642 TTTTGCCATGTTGCGCAGGCTGCTCGAACTCTTGACTTCAGATGATCCTGCTTG 583

QY 1727 GCCTCCACAGTGTGGGATTACAGCGGTGAGCCACCATGCCGCGCTCTTTCTACCTT 1786

Db 592 GCCTCCACAGTGTGGGATTACAGCGGTGAGCCACCATGCCGCGCTCTTTCTACCTT 523

QY 1787 TACACCTGCTTCTTATCTTACCTCAGATCTGTTTTCACACCTTTCATCCCTGCTCT 1846

Db 522 TACACCTGCTTCTTATCTTACCTCAGATCTGTTTTCACACCTTTCATCCCTGCTCT 463

QY 1847 TCACACTGTCTTCCCATGTTTATAGCTGCTTCTTACACTTTGGTTTGAAGGCGAG 1906

Db 462 TCACACTGTCTTCCCATGTTTATAGCTGCTTCTTACACTTTGGTTTGAAGGCGAG 403









GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 08:05:34 ; Search time 218.243 Seconds  
(without alignments)  
7828.876 Million cell updates/sec

Title: US-09-729-674-1  
Perfect score: 3871  
Sequence: 1 ttctctctctctctctctt.....ataaaaaaaaaaaaaaa 3871

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	466.4	12.0	618	3	US-09-328-111-215 Sequence 215, Appl
2	233.6	6.0	43950	4	US-09-735-934A-3 Sequence 3, Appl
3	233.6	6.0	43950	4	US-10-060-332-3 Sequence 3, Appl
4	229.4	5.9	55298	4	US-09-491-356C-1 Sequence 1, Appl
5	229	5.9	111282	4	US-09-754-250-3 Sequence 3, Appl
6	228.8	5.9	9365	4	US-09-608-285A-8 Sequence 8, Appl
7	228.8	5.9	9365	4	US-09-350-836B-8 Sequence 8, Appl
8	228.8	5.9	9365	4	US-09-370-265-8 Sequence 8, Appl
9	228.8	5.9	9365	4	US-09-557-800C-8 Sequence 8, Appl
10	228.8	5.9	14747	4	US-09-608-285A-42 Sequence 42, Appl
11	228.8	5.9	14747	4	US-09-557-800C-42 Sequence 42, Appl
12	228.8	5.9	15977	4	US-09-608-285A-59 Sequence 59, Appl
C 13	228.4	5.9	64467	4	US-09-803-671B-3 Sequence 3, Appl
14	227	5.9	11725	2	US-08-756-506-1 Sequence 1, Appl
C 15	226.6	5.9	55298	4	US-09-491-356C-1 Sequence 1, Appl
16	226.2	5.8	53332	4	US-09-801-861-3 Sequence 3, Appl
17	225.6	5.8	99500	4	US-09-798-096-10 Sequence 10, Appl
C 18	223.8	5.8	16063	4	US-09-801-053-3 Sequence 3, Appl
C 19	223.6	5.8	4773	3	US-08-884-324-9 Sequence 9, Appl
C 20	223.6	5.8	11464	3	US-08-884-324-13 Sequence 13, Appl
21	223.6	5.8	28994	3	US-08-884-324-14 Sequence 14, Appl
22	223.4	5.8	16389	4	US-09-741-154-3 Sequence 3, Appl
23	223	5.8	14581	4	US-08-520-373D-4 Sequence 4, Appl
24	223	5.8	22481	4	US-08-367-841A-43 Sequence 43, Appl
25	223	5.8	22481	5	PCT-US95-07201-43 Sequence 43, Appl
26	223	5.8	22484	4	US-09-875-223-2 Sequence 2, Appl
C 27	223	5.8	23167	4	US-09-499-522-1 Sequence 1, Appl

C 28	222.8	5.8	282	1	US-08-133-629-8 Sequence 8, Appl
C 29	222.6	5.8	1854	4	US-09-620-312D-992 Sequence 992, Appl
C 30	222.2	5.7	70000	4	US-09-851-896-3 Sequence 3, Appl
31	222	5.7	62804	4	US-09-800-960-3 Sequence 3, Appl
C 32	221.8	5.7	15287	4	US-09-817-180-3 Sequence 3, Appl
C 33	221.4	5.7	283	4	US-08-579-445-26 Sequence 26, Appl
C 34	221.2	5.7	20966	4	US-09-984-880-3 Sequence 3, Appl
C 35	221	5.7	63588	4	US-09-873-404-3 Sequence 3, Appl
C 36	221	5.7	92139	4	US-09-918-686-1 Sequence 1, Appl
C 37	220.6	5.7	6139	4	US-08-843-076D-33 Sequence 33, Appl
C 38	220.2	5.7	38564	4	US-09-734-673-3 Sequence 3, Appl
39	220.2	5.7	56516	2	US-08-996-306-1 Sequence 1, Appl
40	220.2	5.7	56516	3	US-09-338-907-1 Sequence 1, Appl
41	220.2	5.7	56516	4	US-09-218-207-1 Sequence 1, Appl
42	220.2	5.7	56520	3	US-09-338-907-179 Sequence 179, Appl
43	220.2	5.7	56520	4	US-09-218-207-179 Sequence 179, Appl
44	220.2	5.7	92139	4	US-09-918-686-1 Sequence 1, Appl
C 45	220	5.7	20674	4	US-09-641-638-651 Sequence 651, Appl

## ALIGNMENTS

RESULT 1  
US-09-328-111-215/c  
; Sequence 215, Application US/09328111  
; Patent No. 6262333  
; GENERAL INFORMATION:  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Derti, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; TITLE OF INVENTION: PRODUCTS  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/328,111  
; EARLIER APPLICATION NUMBER: US 60/088,801  
; EARLIER FILING DATE: 1998-06-10  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 215  
; LENGTH: 618  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(618)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-328-111-215

Query Match	12.0%	Score	466.4	DB	3	Length	618
Best Local Similarity	94.8%	Pred. No.	2.6e-99				
Matches	506	Conservative	0	Mismatches	25	Indels	3
Gaps	3						
Qy	1076	CCAGTACACAGGACACTTGGGGCTCTCGGATGAGCAGCTTCCAGCCAGACCTGAAGC	1135				
Db	534	CNAGNCCNAGNCANTGGGGCTCTNGATGAGCANTCCAGCCAGNCTTAANC	475				
Qy	1136	AGTTCAGTCTTTTAATGTGACCA-CACTCCAGAGGAGC-TGTATGACTTT-GCTAAGGA	1192				
Db	474	AGTTCAGTCTTTAAATGTGNCACCACTCCAGANGAAGCTTGTATGACTTTGGNTAAGGA	415				
Qy	1193	AAATATAATGATGATGATGAGGGAAGTCTGGATATGTGGATGACCTCTTGAAC	1252				

Db 414 AATATATATGATGATGATGAGGAGAGTGTGTGGAATATGTGGATGACCTCTTGAAC 355  
QY 1253 GGAGGAGACAGTACGCCACAGACACAAAGAGACTTCTCTTGGCGTTACAGAAACAC 1312  
Db 354 GGAGGAGACAGTACGCCACAGACACAAAGAGACTTCTCTTGGCGTTACAGAAACAC 295  
QY 1313 AGATTCTTTTCTCTTTCCCAACAGCCAGGCTGTGTACCTCAGAGCTTCTCTTTAC 1372  
Db 294 AGATTCTTTTCTCTTTCCCAACAGCCAGGCTGTGTACCTCAGAGCTTCTCTTTAC 235  
QY 1373 TCTCCAAAGTGAAGGAAGAGCCCGCTCTCTCTTAAGTCAATGTCATCAGGGGTGAGCGTG 1432  
Db 234 TCTCCAAAGTGAAGGAAGAGCCCGCTCTCTCTTAAGTCAATGTCATCAGGGGTGAGCGTG 175  
QY 1433 CTTTCTCTATCTTCACACCTGCCACCTCAATGTTTCACACCTATCTTCTCACCTTTTTTT 1492  
Db 174 CTTTCTCTATCTTCACACCTGCCACCTCAATGTTTCACACCTATCTTCTCACCTTTTTTT 115  
QY 1493 GAGATGAGTCTCGCTCTCTTGGCCAGGCTGGAGTGCATGGAATGTCATGTCAGTCACTG 1552  
Db 114 GAGATGAGTCTCGCTCTCTTGGCCAGGCTGGAGTGCATGGAATGTCATGTCAGTCACTG 55  
QY 1553 CACCTCCGCTCTTGGGTTCAAGCAATCTGCTGCATCAGCTCCCGAGTACC 1606  
Db 54 CACCTCCGCTCTTGGGTTCAAGCAATCTGCTGCATCAGCTCCCGAGTACC 1

## RESULT 2

US-09-735-934A-3  
; Sequence 3, Application US/09735934A  
; Patent No. 6372468  
; GENERAL INFORMATION:  
; APPLICANT: LI, JIAYIN et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CL000851  
; CURRENT APPLICATION NUMBER: US/09/735, 934A  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 43950  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-735-934A-3

Query Match 6.0%; Score 233.6; DB 4; Length 43950;  
Best Local Similarity 86.8%; Pred. No. 4.1e-44;  
Matches 257; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
QY 1478 TCTCACCTTTTTTTGAGATGAGTCTCGCTCTCTTGGCCAGGCTGGAGTGCATGGAAC 1537  
Db 11765 TCTTTTTTTTTTTGAACTGAGTCTCGCTGTATTGCCAGGCTGGAGTGCAGTGGCGT 11824  
QY 1538 GTTCTCAGCTCACTGCAACCTCCGCTCTTGGGTTCAAGCAATCTGCTGCATCAGCCCTC 1597  
Db 11825 GATCTCAGCTCACTGCAACCTCTGCTCTCTGGGTTCAAGTGTATCTCTGCTCAGCCTC 11884  
QY 1598 CCGAGTACCTGGGATACAGGATGTCACCAACCGCCGCTAAATTTGATTTTTAGTA 1657  
Db 11885 CGGAGTACCTGGGATACAGGATGTCACCAACCGCCGCTAAATTTGATTTTTAGTA 11944  
QY 1658 GAGACGGGGTTTGGCAATGTTGCCAGGCTGCTCGAACTCTTGACTTCAGATGATCCA 1717  
Db 11945 GAGATGGGGTTTGGCAATGTTGCCAGGCTGCTCGAACTCTTGACTTCAGATGATCCA 12004  
QY 1718 TCTGCCTTGGCTCCCAACAGTCTGGGATTTACAGGCGTGGAGCCACCATGCCCGGCC 1773  
Db 12005 CCGGCTTGGCTCCCAACAGTCTGGGATTTACAGGCGTGGAGCCACCATGCCCGGCC 12060

## RESULT 3

US-09-735-934A-3

US-10-060-332-3  
; Sequence 3, Application US/10060332  
; Patent No. 6528294  
; GENERAL INFORMATION:  
; APPLICANT: LI, JIAYIN et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CL000851DIV  
; CURRENT APPLICATION NUMBER: US/10/060, 332  
; CURRENT FILING DATE: 2002-02-01  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 43950  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-060-332-3

Query Match 6.0%; Score 233.6; DB 4; Length 43950;  
Best Local Similarity 86.8%; Pred. No. 4.1e-44;  
Matches 257; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
QY 1478 TCTCACCTTTTTTTGAGATGAGTCTCGCTCTCTTGGCCAGGCTGGAGTGCATGGAAC 1537  
Db 11765 TCTTTTTTTTTTTGAACTGAGTCTCGCTGTATTGCCAGGCTGGAGTGCAGTGGCGT 11824  
QY 1538 GTTCTCAGCTCACTGCAACCTCCGCTCTTGGGTTCAAGCAATCTGCTGCATCAGCCTC 1597  
Db 11825 GATCTCAGCTCACTGCAACCTCTGCTCTCTGGGTTCAAGTGTATCTCTGCTCAGCCTC 11884  
QY 1598 CCGAGTACCTGGGATTTACAGGATGTCACCAACCGCCGCTAAATTTGATTTTTAGTA 1657  
Db 11885 CGGAGTACCTGGGATTTACAGGATGTCACCAACCGCCGCTAAATTTGATTTTTAGTA 11944  
QY 1658 GAGACGGGGTTTGGCAATGTTGCCAGGCTGCTCGAACTCTTGACTTCAGATGATCCA 1717  
Db 11945 GAGATGGGGTTTGGCAATGTTGCCAGGCTGCTCGAACTCTTGACTTCAGATGATCCA 12004  
QY 1718 TCTGCCTTGGCTCCCAACAGTCTGGGATTTACAGGCGTGGAGCCACCATGCCCGGCC 1773  
Db 12005 CCGGCTTGGCTCCCAACAGTCTGGGATTTACAGGCGTGGAGCCACCATGCCCGGCC 12060

## RESULT 4

US-09-491-356C-1  
; Sequence 1, Application US/09491356C  
; Patent No. 6566061  
; GENERAL INFORMATION:  
; APPLICANT: Philibert, Robert A.  
; APPLICANT: Ginns, Edward I.  
; APPLICANT: Delisi, Lynn  
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13  
; FILE REFERENCE: 9465.6US11  
; CURRENT APPLICATION NUMBER: US/09/491,356C  
; CURRENT FILING DATE: 2000-01-26  
; PRIOR APPLICATION NUMBER: PCT/US99/09365  
; PRIOR FILING DATE: 1999-04-29  
; PRIOR APPLICATION NUMBER: 60/083,465  
; PRIOR FILING DATE: 1998-04-29  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 55298  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (485)..(485)  
; OTHER INFORMATION: n is not determined  
; NAME/KEY: misc feature  
; LOCATION: (836)..(836)  
; OTHER INFORMATION: n is not determined

```

; NAME/KEY: misc feature
; LOCATION: (16728)..(16728)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc feature
; LOCATION: (22750)..(22750)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc feature
; LOCATION: (22756)..(22756)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc feature
; LOCATION: (28519)..(28519)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc feature
; LOCATION: (44804)..(44804)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc feature
; LOCATION: (45002)..(45002)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc feature
; LOCATION: (54049)..(54049)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc feature
; LOCATION: (54226)..(54226)
; OTHER INFORMATION: n is not determined
; OTHER INFORMATION: n is not determined
US-09-491-356C-1

Query Match      5.9%; Score 229.4; DB 4; Length 55298;
Best Local Similarity 80.7%; Pred. No. 4.4e-43;
Matches 280; Conservative 0; Mismatches 66; Indels 1; Gaps 1;

QY 1471 CTATCTTCTCACCTTTTTTTTGTAGATGGAGTCTCGTCTCTTCCAGGCTGGAGTGCA 1530
Db 42875 CTATTTTTTTTTTTTTTTTGTAGACAGAGTCTCGTCTCTCGCTCAGGCTGGAGTGCA 42934

QY 1531 ATGGACATCTCTAGCTCACTGCAACCTCCCGCTCTTGGGTTCAAGCAATCTCTGCAT 1590
Db 42935 ATGGCATGATCTCGGCTCACTACACCTCCCGCTCCCGGTAAAGCAATCTCTCGCT 42994

QY 1591 CAGCTCCCGAGTACCTGGGATTACAGCATGTGCCACACCGCCGCGC-TAATTTTCTAT 1649
Db 42995 CAGCTCCCGAGTACCTGGGATTACAGCGGCTGCCACACCGCCGCGATTTTTGTAT 43054

QY 1650 TTTTAGTAGACAGCGGGTTTTGGCATGTGTGCCAGGCTGGTCTCGAACTCTTGACTTCAG 1709
Db 43055 TTTTAGTAGACAGCGGGTTTACCATGTGTGCCAGGCTGGTCTTGAACCTCGACCTCAG 43114

QY 1710 ATGATCCATCTGCTTGGCTCCACAGTGTGGATTACAGGCGTGAGCCACCATGCC 1769
Db 43115 GTGATCCACCTGCTCGGCTCCCAAGTGTGGATTACAGGCGATGAGCCATCACGCC 43174

QY 1770 GGCCTCTTCTCACCTTTTACACCTGTCTTCTATCTCTCACATCTGTT 1816
Db 43175 AGCCTCTAAGGTGCTTTTCTAGATATCTTGGTGATTCATGAATGTT 43221

RESULT 5
US-09-754-250-3
; Sequence 3, Application US/09754250
; Patent No. 6376225
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: PHOSPHODIESTERASE PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001063
; CURRENT APPLICATION NUMBER: US/09/754,250
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 111282
; TYPE: DNA
; ORGANISM: Human

; NAME/KEY: misc feature
; LOCATION: (1)..(111282)
; OTHER INFORMATION: n = A,T,C or G
US-09-754-250-3

Query Match      5.9%; Score 229; DB 4; Length 111282;
Best Local Similarity 83.8%; Pred. No. 7.7e-43;
Matches 259; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1480 TCACCTTTTTTTTGTAGATGGAGTCTCGTCTCTTCCAGGCTGGAGTGCAATGCCAGT 1539
Db 106289 TCTTGTTTTTTTTGTAGATGGAGTCTCACTCTGTTCGCCAGGCTGGAGTGCAATGCCAG 106348

QY 1540 TCTCAGCTCACTGCAACCTCCGCTCTTGGGTTCAAGCAATCTCTGATCAGCTCTCC 1599
Db 106349 TCTTGGCTCACTGCAACCTCCGCTCTTGGGTTCAAGCAATCTCTGCTCAGCTCTCC 106408

QY 1600 GAGTACCTGGGATTACAGGCATGTGCCACCGCCGCTAATTTTGTATTTTAGTAGA 1659
Db 106409 AAGTAGCTGGGATTACAGGCATGTGCCACCGCCGCTAATTTTGTATTTTAGTAGA 106468

QY 1660 GACGGGGTTTTCCCATGTGTGGCCAGGCTGGTCTCGAACTCTTGAATCAGATGATCCATC 1719
Db 106469 GATGGGGCTTCACCATTTTGGTCAGGCTGGTCTTGAACCTCTCAGTCAAGTGATCCACA 106528

QY 1720 TGCCTTGGCTCCACAGTGTGGGATTACAGGCGTGAGCCACCATGCCGCGCTCTTTC 1779
Db 106529 CACCTCGGCATCTCGAGTGTGGGATTACAGGCGTGAGCTACCGTACCTGGCCCTTGTG 106588

QY 1780 TCACCTTTA 1788
Db 106589 GAAATTTCA 106597

RESULT 6
US-09-608-285A-8
; Sequence 8, Application US/09608285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: 28110/36570
; CURRENT APPLICATION NUMBER: US/09/608,285A
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/583,231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557,800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 9365
; TYPE: DNA
```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3409)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
; NAME/KEY: misc feature
; LOCATION: (9214)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
; NAME/KEY: misc feature
; LOCATION: (9303)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
; NAME/KEY: misc feature
; LOCATION: (9311)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
US-09-608-285A-8

Query Match
Best Local Similarity 5.9%; Score 228.8; DB 4; Length 9365;
Matches 261; Conservative 34; Mismatches 47; Indels 4; Gaps 2;

QY 1462 TTTTACACCTATCTTTCTCACCTTTTTCAGATGGAGTCTCGTC-TCTTGCCGAG 1520
Db 7413 TATTTATTTATTTATTTATTTATTTTGGAGCGAGTCTGCTCTTGTTCYRGG 7472

QY 1521 CTGGAGTCAATGGCAGCTTCTCAGCTCACTGCACCTCCGCTCTTGGTTCAGCAAT 1580
Db 7473 CTGGAGTCAATGGCAGCTTCTCAGCTCACTGCACCTCCGCTCTTGGTTCAGCAAT 7532

QY 1581 TCTGCTGCATCAGCTCCGAGTACCTGGGATTACAGGATGTCACCAACCCCGGCTA 1640
Db 7533 TCTGCTGCATCAGCTCCGAGTACCTGGGATTACAGGATGTCACCAACCCCGGCTA 7592

QY 1641 A---TTTTGTATTTTAGTAGAGCGGGTTTGGCATGTTGGCCAGCTGCTCGAAC 1697
Db 7593 ATTTTGTATTTTAGTAGAGCGGGTTTGGCATGTTGGCCAGCTGCTCGAAC 7652

QY 1698 TCTGCTGCATCAGCTCCGAGTACCTGGGATTACAGGATGTCACCAACCCCGGCTA 1757
Db 7653 TCTGCTGCATCAGCTCCGAGTACCTGGGATTACAGGATGTCACCAACCCCGGCTA 7712

QY 1758 GCCACATGCGCGGCTCTTTCTACCTTTACACCTGCTCTCTTAT 1803
Db 7713 GCCACATGCGCGGCTCTTTCTACCTTTACACCTGCTCTCTTAT 7758

```

```

RESULT 7
US-09-350-836B-8
; Sequence 8, Application US/09350836B
; Patent No. 6387645
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
; FILE REFERENCE: 28110/35761
; CURRENT APPLICATION NUMBER: US/09/350,836B
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 9365
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon

```

```

; LOCATION: (1)..(288)
; NAME/KEY: exon
; LOCATION: (1281)..(1580)
; NAME/KEY: exon
; LOCATION: (1820)..(1855)
; NAME/KEY: exon
; LOCATION: (2467)..(2555)
; NAME/KEY: exon
; LOCATION: (2863)..(2942)
; NAME/KEY: exon
; LOCATION: (3889)..(3950)
; NAME/KEY: exon
; LOCATION: (4894)..(4995)
; NAME/KEY: exon
; LOCATION: (5847)..(5987)
; NAME/KEY: exon
; LOCATION: (6966)..(7138)
; NAME/KEY: exon
; LOCATION: (8556)..(9365)
; NAME/KEY: misc feature
; LOCATION: (9309)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymine
; NAME/KEY: misc feature
; LOCATION: (9214)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymine
; NAME/KEY: misc feature
; LOCATION: (9303)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymine
; NAME/KEY: misc feature
; LOCATION: (9311)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymine
US-09-350-836B-8

```

```

Query Match
Best Local Similarity 5.9%; Score 228.8; DB 4; Length 9365;
Matches 261; Conservative 34; Mismatches 47; Indels 4; Gaps 2;

QY 1462 TTTTACACCTATCTTTCTCACCTTTTTCAGATGGAGTCTCGTC-TCTTGCCGAG 1520
Db 7413 TATTTATTTATTTATTTATTTATTTTGGAGCGAGTCTGCTCTTGTTCYRGG 7472

QY 1521 CTGGAGTCAATGGCAGCTTCTCAGCTCACTGCACCTCCGCTCTTGGTTCAGCAAT 1580
Db 7473 CTGGAGTCAATGGCAGCTTCTCAGCTCACTGCACCTCCGCTCTTGGTTCAGCAAT 7532

QY 1581 TCTGCTGCATCAGCTCCGAGTACCTGGGATTACAGGATGTCACCAACCCCGGCTA 1640
Db 7533 TCTGCTGCATCAGCTCCGAGTACCTGGGATTACAGGATGTCACCAACCCCGGCTA 7592

QY 1641 A---TTTTGTATTTTAGTAGAGCGGGTTTGGCATGTTGGCCAGCTGCTCGAAC 1697
Db 7593 ATTTTGTATTTTAGTAGAGCGGGTTTGGCATGTTGGCCAGCTGCTCGAAC 7652

QY 1698 TCTGCTGCATCAGCTCCGAGTACCTGGGATTACAGGATGTCACCAACCCCGGCTA 1757
Db 7653 TCTGCTGCATCAGCTCCGAGTACCTGGGATTACAGGATGTCACCAACCCCGGCTA 7712

QY 1758 GCCACATGCGCGGCTCTTTCTACCTTTACACCTGCTCTCTTAT 1803
Db 7713 GCCACATGCGCGGCTCTTTCTACCTTTACACCTGCTCTCTTAT 7758

```

```

RESULT 8
US-09-370-265-8
; Sequence 8, Application US/09370265
; Patent No. 644771
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
; FILE REFERENCE: 28111/35908
; CURRENT APPLICATION NUMBER: US/09/370,265

```





Query Match 5.9%; Score 228.8; DB 4; Length 14747;  
Best Local Similarity 75.4%; Pred. No. 3.1e-43;  
Matches 261; Conservative 34; Mismatches 47; Indels 4; Gaps 2;

QY 1462 TGTTCACACCTATCTTTCTACACCTTTTGTGAGATGAGTCTGCTC-TCTTGCCGAGG 1520  
Db 10773 TATTATTATTATTATTATTATTATTATTATTGTGAGCGAGTCTGCTTTGTTKCYRGG 10832

QY 1521 CTGGAGTGCATGCAAGTCTTCTAGCTCACTGCAACCTCGCTCTTTGGTTCAGCAAT 1580  
Db 10833 CTGGAGTGCATGCAAGTCTTCTAGCTCACTGCAACCTCGCTCTTTGGTTCAGCGAT 10892

QY 1581 TCTGCTGATCAGCTCCCGAGTACCTGGGATACAGGATGCGCCACACGCGCGCTA 1640  
Db 10893 TCTYCTGCTCAGCTCCCGAGTACCTGGGATACAGGATGCGCCACACGCGCGCTA 10952

QY 1641 A---TTTTGTATTATTATTAGTAGAGAGCGGGTTTTCCTATGTCAGGCTGGTCTGAAC 1697  
Db 10953 ATTTTGTATTATTATTAGTAGAGAGCGGGTTTTCCTATGTCAGGCTGGTCTGAAC 11012

QY 1698 TCTGATTCAGATGATCCATCTGCTTGGCTCCACAGTCTGCGATACAGGCGTGA 1757  
Db 11013 TCTGATTCAGATGATCCATCTGCTTGGCTCCACAGTCTGCGATACAGGCGTGA 11072

QY 1758 GCCACCATGCCGCGCTCTTCTACACCTTTACACCTGCTCTTAT 1803  
Db 11073 GCCACCATGCCGCGCTCTTCTACACCTTTACACCTGCTCTTAT 11118

## RESULT 12

US-09-608-285A-59  
; Sequence 59, Application US/09608285A  
; Patent No. 635013  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John  
; APPLICANT: Mulero, Julio  
; APPLICANT: Yeung, George  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE  
; FILE REFERENCE: 28110/36570  
; CURRENT APPLICATION NUMBER: US/09/608,285A  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR FILING DATE: 1999-07-16  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR FILING DATE: 1999-02-04  
; PRIOR FILING DATE: 1999-02-04  
; PRIOR FILING DATE: 1998-07-24  
; PRIOR FILING DATE: 1998-07-16  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 59  
; LENGTH: 15977  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: CD39-L4/L66 Gene Sequence  
; NAME/KEY: CDS  
; LOCATION: (245)..(461)

Query Match 5.9%; Score 228.8; DB 4; Length 15977;  
Best Local Similarity 75.4%; Pred. No. 3.2e-43;  
Matches 261; Conservative 34; Mismatches 47; Indels 4; Gaps 2;

QY 1462 TGTTCACACCTATCTTTCTACACCTTTTGTGAGATGAGTCTGCTC-TCTTGCCGAGG 1520  
Db 12003 TATTATTATTATTATTATTATTATTATTATTGTGAGCGAGTCTGCTTCTGTTKCYRGG 12062

QY 1521 CTGGAGTGCATGCAAGTCTTCTAGCTCACTGCAACCTCGCTCTTTGGTTCAGCAAT 1580  
Db 12063 CTGGAGTGCATGCAAGTCTTCTAGCTCACTGCAACCTCGCTCTTTGGTTCAGCGAT 12122

QY 1581 TCTGCTGATCAGCTCCCGAGTACCTGGGATACAGGATGCGCCACACGCGCGCTA 1640  
Db 12123 TCTYCTGCTCAGCTCCCGAGTACCTGGGATACAGGATGCGCCACACGCGCGCTA 12182

QY 1641 A---TTTTGTATTATTATTAGTAGAGAGCGGGTTTTCCTATGTCAGGCTGGTCTGAAC 1697  
Db 12183 ATTTTGTATTATTATTAGTAGAGAGCGGGTTTTCCTATGTCAGGCTGGTCTGAAC 12242

QY 1698 TCTGATTCAGATGATCCATCTGCTTGGCTCCACAGTCTGCGATACAGGCGTGA 1757  
Db 12243 TCTGATTCAGATGATCCATCTGCTTGGCTCCACAGTCTGCGATACAGGCGTGA 12302

QY 1758 GCCACCATGCCGCGCTCTTCTACACCTTTACACCTGCTCTTAT 1803  
Db 12303 GCCACCATGCCGCGCTCTTCTACACCTTTACACCTGCTCTTAT 12348

## RESULT 13

US-09-803-671B-3/c  
; Sequence 3, Application US/09803671B  
; Patent No. 6582946  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL001161  
; CURRENT APPLICATION NUMBER: US/09/803,671B  
; CURRENT FILING DATE: 2001-03-12  
; NUMBER OF SEQ ID NOS: 7



; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 64467

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(64467)

; OTHER INFORMATION: n = A,T,C or G

US-09-803-671B-3

Query Match 5.9%; Score 228.4; DB 4; Length 64467;

Best Local Similarity 86.1%; Pred. No. 8.1e-43;

Matches 253; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1485 TTTTITTTGAGATGGAGTCTCGCTCTCTTGCCAGGCTGGAGTGAATGGCACGTTCTCA 1544

Db 24074 TTTGTTTITGAGAGCGAGCTCACTCTGTCACCCAGTTGGAGTGAATGGTACAATCTCA 24015

QY 1545 GTCATCTGAACCTCGGCTCTTGGGTTCAAGCAATCTGCTGCATCAGCCTCCCGAGTA 1604

Db 24014 GTCATCTGAACCTCGGCTCTTGGGTTCAAGCAATCTGCTGCATCAGCCTCCCGAGTA 23955

QY 1605 CTGGGATTACAGGATGTCACACGCGCGGCTAAATTTGTATTTTGTAGTAGAGCGG 1664

Db 23954 GTTGGATTACAGTGGCGACACACCGGCTAAATTTGTATTTTGTAGTAGAGCGG 23895

QY 1665 GTTTGTGCATGTTGGCAGGCTGTCGTGGAATCTTGACTTCAGATCAATCTCGCT 1724

Db 23894 GTTTGTGCATGTTGGCAGGCTGTCGTGGAATCTTGACTTCAGATCAATCTCGCT 23835

QY 1725 TGCGCTCCACAGTCTGGATTACAGCGTGAGCCACCATGCCCGGCTCTTT 1778

Db 23834 TGCGCTCCACAGTCTGGATTACAGCGTGAGCCACCATGCCCGGCTCTTT 23781

RESULT 14

US-08-756-506-1

; Sequence 1, Application US/08756506

; Patent No. 5905185

; GENERAL INFORMATION:

; APPLICANT: Garner, Ian

; APPLICANT: Cottingham, Ian R.

; APPLICANT: Temperley, Simon M.

; APPLICANT: Foster, Donald C.

; APPLICANT: Sprecher, Cindy A.

; APPLICANT: Frankard, Donna E.

; TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ZymoGenetics, Inc.

; STREET: 1201 Eastlake Avenue East

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98102

; COMPUTER READABLE FORM:

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/756,506

; FILING DATE:

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Sawislak, Deborah A.

; REGISTRATION NUMBER: 37,438

; REFERENCE/DOCKET NUMBER: 95-28

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-442-6672

; TELEFAX: 206-442-6678

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 11725 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: Join(3520..3630, 5093..5117, 5210..5347, 5450

; LOCATION: ..5584, 8253..8395, 9269..9386, 10516..11102)

US-08-756-506-1

Query Match

Best Local Similarity 85.0%; Pred. No. 7.2e-43;

Matches 266; Conservative 0; Mismatches 45; Indels 2; Gaps 1;

QY 1469 ACCTATCTTTCTCACCTTTTTTTTGAGATGGAGTCTCGCTCT--CTTGCCAGGCTGGAG 1526

Db 5361 ACCTCTTTTTTTCTTTTTTTTGAGATGGAGTTCACCTCTGTGTTGCCAGGCTGGAG 5420

QY 1527 TGAATGGCACGTTCTCAGCTCACTGCAACCTCCGCTCTTGGGTTCAAGCAATCTGCT 1586

Db 6421 TGAATGACGTGATCTCAGCTCACCACAACTCCGCTCTCGCTCTTCAAGCGATCTCTCT 6480

QY 1587 GCATCAGCTCCGAGTACCTGGATTACAGCATCTGCACACGCGGCTAAATTTTG 1646

Db 6481 GCCGAGCTCCGAGTACCTGGATTACAGCATCTGCACACGCGGCTAAATTTTG 5540

QY 1647 TATTTTATAGTAGAGAGCGGGTTTTTGCATGTTGGCAGGCTGCTCTCGAACTCTTGA 1706

Db 6541 TGTTTTTAGTAGAGAGAGGGTTTTCTCCGTTGTTGTTCAAGCTGCTCTTGA 6600

QY 1707 CAGATGATCAATCTGCTTGGCTCCACAGTCTGGGATTACAGCGTGAAGCGTGAAGCG 1766

Db 6601 CAGTGTATCACTGCTTGGCTCCACAGTCTGGGATTACAGCGTGAAGCGTGAAGCG 6660

QY 1767 CCGGCGCTCTTTTC 1779

Db 6661 CCCAGCTCTTTTC 6673

RESULT 15

US-09-491-356C-1/c

; Sequence 1, Application US/09491356C

; Patent No. 6566061

; GENERAL INFORMATION:

; APPLICANT: Philibert, Robert A.

; APPLICANT: Gibbs, Edward I.

; APPLICANT: Delisi, Lynn

; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13

; FILE REFERENCE: 9465.6US11

; CURRENT APPLICATION NUMBER: US/09/491,356C

; CURRENT FILING DATE: 2000-01-26

; PRIOR APPLICATION NUMBER: PCT/US99/09365

; PRIOR FILING DATE: 1999-04-29

; PRIOR APPLICATION NUMBER: 60/083,465

; PRIOR FILING DATE: 1998-04-29

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 1

; LENGTH: 55298

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (485)..(485)

; OTHER INFORMATION: n is not determined

; NAME/KEY: misc feature

; LOCATION: (838)..(838)

; OTHER INFORMATION: n is not determined

; NAME/KEY: misc feature

; LOCATION: (16728)..(16728)

OTHER INFORMATION: n is not determined  
NAME/KEY: misc feature  
LOCATION: (22750)..(22750)  
OTHER INFORMATION: n is not determined  
NAME/KEY: misc feature  
LOCATION: (22756)..(22756)  
OTHER INFORMATION: n is not determined  
NAME/KEY: misc feature  
LOCATION: (28519)..(28519)  
OTHER INFORMATION: n is not determined  
NAME/KEY: misc feature  
LOCATION: (44804)..(44804)  
OTHER INFORMATION: n is not determined  
NAME/KEY: misc feature  
LOCATION: (45002)..(45002)  
OTHER INFORMATION: n is not determined  
NAME/KEY: misc feature  
LOCATION: (54049)..(54049)  
OTHER INFORMATION: n is not determined  
NAME/KEY: misc feature  
LOCATION: (54226)..(54226)  
OTHER INFORMATION: n is not determined

US-09-491-356C-1

Query Match 5.9%; Score 226.6; DB 4; Length 55298;  
Best Local Similarity 83.8%; Pred. No. 2e-42;  
Matches 269; Conservative 0; Mismatches 49; Indels 3; Gaps 1;

Qy	1471	CTATCTTCACACCTTTTGTGAGATGAGTCTCGCTCTCTGCCCCAGGCTGGAGTGCA	1530
Db	33384	CTTTTCTTTTCTTTTGTGAGATGAGTCTCGCTCTCTGCCCCAGGCTGGAGTGCA	33325
Qy	1531	ATGGCAGGTTCTCAGCTCACTCAACCTCCGCTCTTGGGTTCAAGCAATTCGTGTCAT	1590
Db	33324	GTGGCAGATCTCGCTCACTCAAGCTCTGCTCTCTGGGTTATGCAATTCCTGCTCT	33265
Qy	1591	CAGCTCCCGAGTACCTGGGATTAAGGATGTCGCCACCCAGCCCGGCTAATTTGT---	1647
Db	33264	CAACCTCCCGAGTACCGGAACTACAGGCACCCCGCCACCCAGCTAATTTTGTGA	33205
Qy	1648	ATTTTGTAGAGACGGGTTTTCCTATGTTGGCCAGGCTGGTCTCGAATCTTGACTTC	1707
Db	33204	TTTTTGTAGAGACGGGTTTTCCTATGTTGGCCAGGATGGTCTCGAATCTCGACTTC	33145
Qy	1708	AGATGATCCATCTGCTTGGCTCCCAAGTGTGGGATTAAGGCTGAGCCACCATGC	1767
Db	33144	AGGTGATCCACCTGCTCGGCTCCCAAGTGTGGGATTAAGGCTGAGCCACCATGC	33085
Qy	1768	CCGGCTCTTTCTCACCTTTA	1788
Db	33084	CCGGCTCTCTCTGAGGTTA	33064

Search completed: January 30, 2004, 18:32:22  
Job time : 222.243 secs





```
QY 2221 CCATCCCAATTGAGGCAATCTTAGTTAATAGTCTCCAGTTTGAAAAGGAAGGC 2280
Db 2221 CCATCCCAATTGAGGCAATCTTAGTTAATAGTCTCCAGTTTGAAAAGGAAGGC 2280
QY 2281 CAGAGCTGTCTAGTATTCATTCTTTTCAGTAAATATTTATTCAGTACCTACTGTGTGC 2340
Db 2281 CAGAGCTGTCTAGTATTCATTCTTTTCAGTAAATATTTATTCAGTACCTACTGTGTGC 2340
QY 2341 TAGGCATTGACCTGGGAATAGATATCTTCACAGAAATACAGGGAAGTTCCCTGTGCT 2400
Db 2341 TAGGCATTGACCTGGGAATAGATATCTTCACAGAAATACAGGGAAGTTCCCTGTGCT 2400
QY 2401 CATGGAGCTTACATTTACAGGGAAGAGATAGCCAAATACATAGGAATATAATATATAC 2460
Db 2401 CATGGAGCTTACATTTACAGGGAAGAGATAGCCAAATACATAGGAATATAATATATAC 2460
QY 2461 AAGGTATCATGTAGTATTAATGCTGTGGAGAAAATAAGCAGGGGAGGAGTAAGAAA 2520
Db 2461 AAGGTATCATGTAGTATTAATGCTGTGGAGAAAATAAGCAGGGGAGGAGTAAGAAA 2520
QY 2521 TCCTGGAGATGAGCTGCAGTTTTAAATGGGGCCCTCACTGGGAATGTGACGTTGAGCAGA 2580
Db 2521 TCCTGGAGATGAGCTGCAGTTTTAAATGGGGCCCTCACTGGGAATGTGACGTTGAGCAGA 2580
QY 2581 GACGTTAGGGAAGTGGATCCCTGGACAAGGCATTCAGGCAGAGGAACAAGATGTCACTG 2640
Db 2581 GACGTTAGGGAAGTGGATCCCTGGACAAGGCATTCAGGCAGAGGAACAAGATGTCACTG 2640
QY 2641 CCCCAGAGTGAAGACTTGCTCTAGCTGTGTGAGAAAAGAGCAGGAGACCAAGCAGAGTCG 2700
Db 2641 CCCCAGAGTGAAGACTTGCTCTAGCTGTGTGAGAAAAGAGCAGGAGACCAAGCAGAGTCG 2700
QY 2701 TGGGAGGGGTAGAAATGGAAGAGGAGCGGCTGGGAGGACAGGTGTGGAGGCGCTTGG 2760
Db 2701 TGGGAGGGGTAGAAATGGAAGAGGAGCGGCTGGGAGGACAGGTGTGGAGGCGCTTGG 2760
QY 2761 CTCTGCTTAAGTGAAGTGGGAACCACTGAGAGGTTTGAACAGAGAGGTGCTTGATTGAT 2820
Db 2761 CTCTGCTTAAGTGAAGTGGGAACCACTGAGAGGTTTGAACAGAGAGGTGCTTGATTGAT 2820
QY 2821 TTATATTTGCAAGGTCATCTAGTGCATATTTGTGAAAACCTTTAGTGGACAAGGCG 2880
Db 2821 TTATATTTGCAAGGTCATCTAGTGCATATTTGTGAAAACCTTTAGTGGACAAGGCG 2880
QY 2881 AGAAGGAGGAGGAGACCTGTTAGGAGCTACTGCAAGGTTCCAGGCTTGGGCGCTGGC 2940
Db 2881 AGAAGGAGGAGGAGACCTGTTAGGAGCTACTGCAAGGTTCCAGGCTTGGGCGCTGGC 2940
QY 2941 CACAGCAACAGCAGTGGTCAATATCTAGATTTATTTTGAAGAGCCTAGGATTTGC 3000
Db 2941 CACAGCAACAGCAGTGGTCAATATCTAGATTTATTTTGAAGAGCCTAGGATTTGC 3000
QY 3001 TGAGAGTTGAAATGGAGTGTAGAGAGGAAAGATTAATGATGACATTAAGGTTTTTG 3060
Db 3001 TGAGAGTTGAAATGGAGTGTAGAGAGGAAAGATTAATGATGACATTAAGGTTTTTG 3060
QY 3061 GCCTGAATAGCAGGAAAGATGAGTTACCACTTACTGAAATAGGGAAGGATGGGCTGGT 3120
Db 3061 GCCTGAATAGCAGGAAAGATGAGTTACCACTTACTGAAATAGGGAAGGATGGGCTGGT 3120
QY 3121 AAGTAAGGAAATTTGGTCAAGAGCAGGCTGTCTGGTTTGAATGGGAGGTTCTGGCTGCA 3180
Db 3121 AAGTAAGGAAATTTGGTCAAGAGCAGGCTGTCTGGTTTGAATGGGAGGTTCTGGCTGCA 3180
QY 3181 AATCAAGTGGAGATTTCTCTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGG 3240
Db 3181 AATCAAGTGGAGATTTCTCTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGG 3240
QY 3241 TGCACCTTGGTTATTGTTGGGGTGTCTCTCAGAGGAACTGTGTAAGGCTTTATCAGTC 3300
Db 3241 TGCACCTTGGTTATTGTTGGGGTGTCTCTCAGAGGAACTGTGTAAGGCTTTATCAGTC 3300
```

```
QY 3301 ATTTATGGCTGTGAGAAATTTCTCTGGAGTGTGGGTACATTTGAAGCAAGTACTTCA 3360
Db 3301 ATTTATGGCTGTGAGAAATTTCTCTGGAGTGTGGGTACATTTGAAGCAAGTACTTCA 3360
QY 3361 GTTGAGGCAAGTCTCTGGAAGAGGCTGTAGGATCTGGCAGCTACCATCGTGGTAG 3420
Db 3361 GTTGAGGCAAGTCTCTGGAAGAGGCTGTAGGATCTGGCAGCTACCATCGTGGTAG 3420
QY 3421 TGTGTTGGGGGTGGGGGTCTCTGGGCACTGGCTGTGTGAAGGATCTGGCAGGCAACCACA 3480
Db 3421 TGTGTTGGGGGTGGGGGTCTCTGGGCACTGGCTGTGTGAAGGATCTGGCAGGCAACCACA 3480
QY 3481 GGGCCCCCTACTGAACCAATCAGCATGTCACTGGCATTTAAAGCCATGAGCTGGAGGGC 3540
Db 3481 GGGCCCCCTACTGAACCAATCAGCATGTCACTGGCATTTAAAGCCATGAGCTGGAGGGC 3540
QY 3541 CACTGAGATGTCTCTGAGTATTACTGAGAAACACAGAAAGAGCCATGGATGAGGCC 3600
Db 3541 CACTGAGATGTCTCTGAGTATTACTGAGAAACACAGAAAGAGCCATGGATGAGGCC 3600
QY 3601 TTGGGCTCTCTGGAAATGGGAAATCAGCCAAAGCACTGAGAGGATTAACCTTAAGGTC 3660
Db 3601 TTGGGCTCTCTGGAAATGGGAAATCAGCCAAAGCACTGAGAGGATTAACCTTAAGGTC 3660
QY 3661 AGAGAAAACCAAGAGAGTGTGTGTTCTGGAAGCTGAGCTTTCTTTATTTCAACCTCATTC 3720
Db 3661 AGAGAAAACCAAGAGAGTGTGTGTTCTGGAAGCTGAGCTTTCTTTATTTCAACCTCATTC 3720
QY 3721 CCTTCTCCAAATTAAGCCACTTTGTGTAGTTGGGCCCCCTCCAGGTTTGAAGGCAAGAGAGA 3780
Db 3721 CCTTCTCCAAATTAAGCCACTTTGTGTAGTTGGGCCCCCTCCAGGTTTGAAGGCAAGAGAGA 3780
QY 3781 AAGGCACAGCTTTGGGAAACAGAGACTTTCTCTGCAATAGCTGGGAAGGATTAAGAGA 3840
Db 3781 AAGGCACAGCTTTGGGAAACAGAGACTTTCTCTGCAATAGCTGGGAAGGATTAAGAGA 3840
QY 3841 TAGAGTGTATAAATAAAAAAATAAAAAA 3871
Db 3841 TAGAGTGTATAAATAAAAAAATAAAAAA 3871
```

## RESULT 2

```
US-09-925-300-660
; Sequence 660, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 660
; LENGTH: 2549
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-660
```

Query Match 53.0%; Score 2051.2; DB 10; Length 2549;

Best Local Similarity 99.2%; Pred No. 0;  
Matches 2077; Conservative 6; Mismatches 4; Indels 7; Gaps 2;

```
QY 1763 CATGCCCGGCTCTTTCTCACCTTTACACCTGTCTTTATTCCTCACATCTGTTTCA 1822
Db 206 CTGCGCGGCTCTTTCTCACCTTTACACCTGTCTTTATTCCTCACATCTGTTTCA 265
QY 1823 CCTTCATCCCTCTCTCTCTCATGTTTCAACATTTGTTCTCCCATGTTTCATAGCTGCTTTC 1882
```

Db 266 CTTTCATCCCTGCTCTCTCATGTTACACATTTGCTCTCCCATGTTTCATAGTGCCTTTC 325  
QY 1883 TTACCAATTTTGGTTGAAGGCGAGTCTCTCTGGCTTGTCTTTTGTCTTTTCCCGAAGAAA 1942  
Db 326 TTACCAATTTTGGTTGAAGGCGAGTCTCTCTGGCTTGTCTTTTGTCTTTTCCCGAAGAAA 385  
QY 1943 TCAGTATTTATTTTAAATPAAGAAAACATTTCTAGAGATGATATTTGTGAAACCTTCC 2002  
Db 386 TCAGTATTTATTTTAAATPAAGAAAACATTTCTAGAGATGATATTTGTGAAACCTTCC 445  
QY 2003 TTTGGCTTATTTGCTTTTCCAGATTTTATGTTCTCTCTTTTCCCATCCCGGAAAAGATGTTG 2062  
Db 446 TTTGGCTTATTTGCTTTTCCAGATTTTATGTTCTCTCTTTTCCCATCCCGGAAAAGATGTTG 505  
QY 2063 GAAGACATAGGCTAAATTTCTCCAGCTCCACAAATGTTCTTCACTTGTCTGACTTGTACC 2122  
Db 506 GAAGACATAGGCTAAATTTCTCCAGCTCCACAAATGTTCTTCACTTGTCTGACTTGTACC 565  
QY 2123 AATTCCTAGACCCACTTGAAAAACAAAGTTGAGTAGAGAGTGTAGAGTGCAGAAAATGTGCT 2182  
Db 566 AATTCCTAGACCCACTTGAAAAACAAAGTTGAGTAGAGAGTGTAGAGTGCAGAAAATGTGCT 625  
QY 2183 TTTGCCCACTTTGCAATCTCCAAATTTACAAAGTTTGGCGATCCCATTTGAGGACAATG 2242  
Db 626 TTTGCCCACTTTGCAATCTCCAAATTTACAAAGTTTGGCGATCCCATTTGAGGACAATG 685  
QY 2243 CTTAGTTATTAAGTCTCCGAGTTGGAAAAGGAAAGCCAGAGCTGTCTAGTTTCATTCA 2302  
Db 686 CTTAGTTATTAAGTCTCCGAGTTGGAAAAGGAAAGCCAGAGCTGTCTAGTTTCATTCA 745  
QY 2303 TTTCTTCAGTAATATTTTATTAAGTACCTTACTGTGTGTAGGATGACCTGG-----G 2356  
Db 746 TTTCTTCAGTAATATTTTATTAAGTACCTTACTGTGTGTAGGATGACCTGGGAAGTAG 805  
QY 2357 AACTAGAGTACTTTCACAGATAAAGGAGGAGTTCCCTGTCTCATGGAGCTTACATTC 2416  
Db 806 AACTAGAGTACTTTCACAGATAAAGGAGGAGTTCCCTGTCTCATGGAGCTTACATTC 865  
QY 2417 TACAGGAGAAAGAGATAGCCAAATACATAGGAATAAATATATACAAAGTATCATGTAGTG 2476  
Db 866 TACAGGAGAAAGAGATAGCCAAATACATAGGAATAAATATATACAAAGTATCATGTAGTG 925  
QY 2477 ATAAATTTCTGTGAGAAAATAAAGCAGGAGGAGTAAGAAATCCCTGGAGATGAGCT 2536  
Db 926 ATAAATTTCTGTGAGAAAATAAAGCAGGAGGAGTAAGAAATCCCTGGAGATGAGCT 985  
QY 2537 GCAGTTTAAATGGGCTTCACTGGGAATGTGACGTTGAGCAGAGACGTTAGGGAAGTGG 2596  
Db 986 GCAGTTTAAATGGGCTTCACTGGGAATGTGACGTTGAGCAGAGACGTTAGGGAAGTGG 1045  
QY 2597 ATCTGACAGAGCATTCAGGCGAGGAAACAGATGTGCACTGCCCCCAAGTGAGAACT 2656  
Db 1046 ATCTGACAGAGCATTCAGGCGAGGAAACAGATGTGCACTGCCCCCAAGTGAGAACT 1105  
QY 2657 TGCTCTACGTGTGTCAGAAAAGACAGGAGACCAAGCAGAGTCGTGGGAGGAGTGAAT 2716  
Db 1106 TGCTCTACGTGTGTCAGAAAAGACAGGAGACCAAGCAGAGTCGTGGGAGGAGTGAAT 1165  
QY 2717 GGAAGAGAGCGCGCTGGGAGAGACAGGTTGTTGGAGGCGCTTGGCTTTCTGCTAAGTGAGA 2776  
Db 1166 GGAAGAGAGCGCGCTGGGAGAGACAGGTTGTTGGAGGCGCTTGGCTTTCTGCTAAGTGAGA 1225  
QY 2777 TGGGAACCACTGAGGAGGTTGAACAGAGGAGTGCCTTGTATTTATTTTGAAGGG 2836  
Db 1226 TGGGAACCACTGAGGAGGTTGAACAGAGGAGTGCCTTGTATTTATTTTGAAGGG 1285  
QY 2837 TCATTTCTAGTGCATATTTGTGAAAACTTTAGTGACAAGGCGACAGGAGAGGGAAG 2896  
Db 1286 TCATTTCTAGTGCATATTTGTGAAAACTTTAGTGACAAGGCGACAGGAGAGGGAAG 1345  
QY 2897 ACCTGTTAGGAACCTACTGCAAGTTCCAGGCTTGGGCTTGGGCTGCGACACACAGAGTG 2956

Db 1346 ACCTGTTAGGAAGCTACTGCAAGGTTCCAGGCTTGGGCTTGGGCCACAGCAACAGCAGTG 1405  
QY 2957 GTCAATATCTAGATTTATTTTGAAGAGCCAAATAGGATTTTCTGAGAGTTTGAATGTG 3016  
Db 1406 GTCAATATCTAGATTTATTTTGAAGAGCCAAATAGGATTTTCTGAGAGTTTGAATGTG 1465  
QY 3017 GAGTGAAGAGAGGAGAGTAAATGATGACATTAAGGTTTTTGGCCTCAATAGCAGGAA 3076  
Db 1466 GAGTGAAGAGAGGAGAGTAAATGATGACATTAAGGTTTTTGGCCTCAATAGCAGGAA 1525  
QY 3077 AGATGAGAGTTCCAGTTACTGAAATAGGAGGATGGCTGGTAAATAGGATTTGCT 3136  
Db 1526 AGATGAGAGTTCCAGTTACTGAAATAGGAGGATGGCTGGTAAATAGGATTTGCT 1585  
QY 3137 GCAAAGCAGGCTGCTGTGTGGTGGAAATGGAGGTTCTGGCTGCAAAATCAAAAGTGAGAA-T 3195  
Db 1586 GCAAAGCAGGCTGCTGTGTGGTGGAAATGGAGGTTCTGGCTGCAAAATCAAAAGTGAGAGT 1645  
QY 3196 TCTCTCAGGTCAGGTCGAGCAGAGCTCGAGACAGGATCTGAATGCACTTGGTTTAT 3255  
Db 1646 TCTCTCAGGTCAGGTCGAGCAGAGCTCGAGACAGGATCTGAATGCACTTGGTTTAT 1705  
QY 3256 GTTGGGGTCTCTCAGAGGACCTGTGAAAGCCCTTTATCAGTCAATTTATTTGGCTGTGA 3315  
Db 1706 GTTGGGGTCTCTCAGAGGAACTGTGAAAGCCCTTTATCAGTCAATTTATTTGGCTGTGA 1765  
QY 3316 GAAATTTCTTGGAGTGTGGGTACATTTGAAGCAAGTGAATTCAGTCAAGGCAAGTCT 3375  
Db 1766 GAAATTTCTTGGAGTGTGGGTACATTTGAAGCAAGTGAATTCAGTCAAGGCAAGTCT 1825  
QY 3376 CTGAAAAGAGGCTGAGGATCTGGCAGCTACCATCGTGGTGTAGTGTGTGGGGTGG 3435  
Db 1826 CTGAAAAGAGGCTGAGGATCTGGCAGCTACCATCGTGGTGTAGTGTGTGGGGTGG 1885  
QY 3436 GGTCTCTGGCAGTGGTGTGTGAAGGATCTGGCAGGCAACCAAGGCGCCCTTACTGAA 3495  
Db 1886 GGTCTCTGGCAGTGGTGTGTGAAGGATCTGGCAGGCAACCAAGGCGCCCTTACTGAA 1945  
QY 3496 CCATCAGCATGTCAAGTGGCATTTAAAGCCATGAGCTGGAGGGCCACTGAGATTTCTC 3555  
Db 1946 CCATCAGCATGTCAAGTGGCATTTAAAGCCATGAGCTGGAGGGCCACTGAGATTTCTC 2005  
QY 3556 TGAGTATTACTGAGAACCAAGAGCCATGGATGAGGCTTGGCTCTCTGGGA 3615  
Db 2006 TGAGTATTACTGAGAACCAAGAGCCATGGATGAGGCTTGGCTCTCTGGGA 2065  
QY 3616 AATGGGAATCAGCCAAAGACTGAGAGAGTTACCTTAAGTCAAGAAAACCAAGAG 3675  
Db 2066 AATGGGAATCAGCCAAAGACTGAGAGAGTTACCTTAAGTCAAGAAAACCAAGAG 2125  
QY 3676 AGTGTGTGTCTGGAGCTGAGCTTTCTTATTCAACCTCATTCCTTCTCCAAATAAG 3735  
Db 2126 AGTGTGTGTCTGGAGCTGAGCTTTCTTATTCAACCTCATTCCTTCTCCAAATAAG 2185  
QY 3736 CCATCTGTAGTTGGGCGCTCCAGGTTGAAGCAAGAGGAGAAAGCAGAGCTTTG 3795  
Db 2186 CCATCTGTAGTTGGGCGCTCCAGGTTGAAGCAAGAGGAGAAAGCAGAGCTTTG 2245  
QY 3796 GGAACAGAGCTTTCTGCAATAGCTGGAGGATTAAGGATAGAGTGT 3849  
Db 2246 GGAACAGAGCTTTCTGCAATAGCTGGAGGATTAAGGATAGAGTGT 2299

## RESULT 3

US-10-108-260A-2058  
; Sequence 2058, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108, 260A  
; CURRENT FILING DATE: 2002-03-27

```

; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1.
; SEQ ID NO 2058
; LENGTH: 1810
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-2058

Query Match      42.8%; Score 1656; DB 12; Length 1810;
Best Local Similarity 93.1%; Pred. No. 0;
Matches 1806; Conservative 2; Mismatches 2; Indels 130; Gaps 2;

Qy 39 CTTTCTCTCTTTCCTCCGGGCGCGATGAGACCGGGGCGCGGGGCGCGGGCGCTGCTA 98
Db 1 CTTTCTCTCTTTCCTCCGGGCGCGATGAGACCGGGGCGCGGGGCGCGGGCGCTGCTA 60
Qy 99 GGGTGTCTGTGGTGGCTGTGGCGCTGTGGCGCGCGGGGCGCGCCCAATACGAACGCTACAGC 158
Db 61 GGGTGTCTGTGGTGGCTGTGGCGCTGTGGCGCGCGGGGCGCGCCCAATACGAACGCTACAGC 120
Qy 159 TTCCGACGCTTCCACCGGAGAGAGCTGATGCGGCTCGAGTCTGGGCTACCGGACGCGCTG 218
Db 121 TTCCGACGCTTCCACCGGAGAGAGCTGATGCGGCTCGAGTCTGGGCTACCGGACGCGCTG 180
Qy 219 GACAGTACAGCGGAGAGCTGGGCGGAGAGCTGGGCTACCTGAGATCAGGCTGGG 278
Db 181 GACAGTACAGCGGAGAGCTGGGCGGAGAGCTGGGCTACCTGAGATCAGGCTGGG 240
Qy 279 CTGCAACCGCTGTCTCGGCGAGCAGAGAGCGCTTCTGCCACCGCAACTGCGAGCGCGCGCG 338
Db 241 CTGCAACCGCTGTCTCGGCGAGCAGAGAGCGCTTCTGCCACCGCAACTGCGAGCGCGCGCG 300
Qy 339 CAGCCGAGCGCGCGCGCGCTCGGCGAGCTATCCCGAGCTGCGGCTTTCGGGGGCGTG 398
Db 301 CAGCCGAGCGCGCGCGCGCTCGGCGAGCTATCCCGAGCTGCGGCTTTCGGGGGCGTG 360
Qy 399 CTGCGCGCGCGCACTGTGCTCAAGCGCTGCAAGCAGAGGCGCTGCCAGCGCTTCCGCGAGTCC 458
Db 361 CTGCGCGCGCGCACTGTGCTCAAGCGCTGCAAGCAGAGGCGCTGCCAGCGCTTCCGCGAGTCC 420
Qy 459 CAGCCGAGCGGAGGCTGTGGCGAGCTTCAGCGCGCGGAGCGCTCAAGGTTCTCGAG 518
Db 421 CAGCCGAGCGGAGGCTGTGGCGAGCTTCAGCGCGCGGAGCGCTCAAGGTTCTCGAG 480
Qy 519 TTCGTTACTTCAAGGCAAAATATCTCCCAAGCCATCGCGCTGCTCACACCTTTCTA 578
Db 481 TTCGTTACTTCAAGGCAAAATATCTCCCAAGCCATCGCGCTGCTCACACCTTTCTA 540
Qy 579 CTGAAGCATCCTGATGACGAAATGATGAAGAGGAACATGGCATATTATAGAGCGCTGCCT 638
Db 541 CTGAAGCATCCTGATGACGAAATGATGAAGAGGAACATGGCATATTATAGAGCGCTGCCT 600
Qy 639 GGTGCGGAGGACTACATTAAAGACCTGGAACCAAGTCAATGAAGCGCTGTTTCATCCGA 698
Db 601 GGTGCGGAGGACTAC-TTAAAGACCTGGAACCAAGTCAATGAAGCGCTGTTTCATCCGA 659
Qy 699 GCAGTGTGGGATACAAACGGTGTGAGAACTGTGAGAAACATCCATCACAGACATGGAGCTGGCC 758
Db 660 GCAGTGTGGGATACAAACGGTGTGAGAACTGTGAGAAACATCCATCACAGACATGGAGCTGGCC 719
Qy 759 CTTCCGAGCTTCTTCAAGGCTTTTACAGTGTCTCGAGCGCTCGAGGGGTTCCAGGAG 818
Db 720 CTTCCGAGCTTCTTCAAGGCTTTTACAGTGTCTCGAGCGCTCGAGGGGTTCCAGGAG 779
Qy 819 ATCAAGGACTTCAAGGATTTCTACCTTCCATAGCAGATCATTTATGATAGAGTTCTCGAA 878
Db 780 ATCAAGGACTTCAAGGATTTCTACCTTCCATAGCAG----- 816
Qy 879 TGAATAATACAGTGTGAAGAGAACTCACCCGAGTTATAGAGGCTATCCGGTTTGAGAAA 938
Db 817 ----- 816
Qy 939 TTGTGGCTACCATGATCATCTACTTGTGAGTTTGTGCTATTATAGTTGAACGACCTGAAG 998
```

RESULT 4



US-10-242-535A-37814  
; Sequence 37814, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: Chondrogene Inc.  
; APPLICANT: Liaw, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 37814  
; LENGTH: 480  
; TYPE: DNA  
; ORGANISM: Human  
US-10-242-535A-37814

Query Match 12.4%; Score 480; DB 12; Length 480;  
Best Local Similarity 100.0%; Pred. No. 1.1e-119;  
Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1377 CAAAGTGAAGGAGGAGCCCGCTCTCTAACTGATGTCATCAGGGGTGACCTGCCTT 1436  
Db 1 CAAAGTGAAGGAGGAGCCCGCTCTCTAACTGATGTCATCAGGGGTGACCTGCCTT 60  
QY 1437 TCCTATCTTCACACCTGCCACCTCATGTTCAACCTATCTTCTCACTTTTGTAGA 1496  
Db 61 TCCTATCTTCACACCTGCCACCTCATGTTCAACCTATCTTCTCACTTTTGTAGA 120  
QY 1497 TGGAGTCTCGCTCTCTGCGGCTGAGTGCATGGAAGTCTCAGCTCAGCTGCAAC 1556  
Db 121 TGGAGTCTCGCTCTCTGCGGCTGAGTGCATGGAAGTCTCAGCTCAGCTGCAAC 180  
QY 1557 CTCGGCTCTTGGGTTCAAGCAATCTGCTGATCAGCTCCCGAGTACCTGGGATTACA 1616  
Db 181 CTCGGCTCTTGGGTTCAAGCAATCTGCTGATCAGCTCCCGAGTACCTGGGATTACA 240  
QY 1617 GGCATGTGCCACACCGCGCTAATTTTGTATTTTAGTAGAGCGGGTTTGGCATG 1676  
Db 241 GGCATGTGCCACACCGCGCTAATTTTGTATTTTAGTAGAGCGGGTTTGGCATG 300  
QY 1677 TTGGCCAGGCTGTCTCGAACTCTTGACTTCAGATGATCCATCTGGCTTGGCTCCCA 1736  
Db 301 TTGGCCAGGCTGTCTCGAACTCTTGACTTCAGATGATCCATCTGGCTTGGCTCCCA 360  
QY 1737 GTGCTGGATTACAGCGGTGAGCCACATGCGCGCTCTTCTCACTTTACACTGTGTC 1796  
Db 361 GTGCTGGATTACAGCGGTGAGCCACATGCGCGCTCTTCTCACTTTACACTGTGTC 420  
QY 1797 TTCTTATCTTCACATCTGTTTTCACACCTTCATCTCCCTGCTTCTCCTCATGTTT 1856  
Db 421 TTCTTATCTTCACATCTGTTTTCACACCTTCATCTCCCTGCTTCTCCTCATGTTT 480

## RESULT 5

US-10-094-749-724  
; Sequence 724, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI

APPLICANT: ISHII, SHIZUKO  
APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUUKO  
APPLICANT: HIO, YURI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMECHIKA, ICHIRO  
APPLICANT: SEKI, NAOHICO  
APPLICANT: YOSHIKAWA, TSUTOMU  
APPLICANT: OTSUKA, MOTOYUKI  
APPLICANT: NAGAHARI, KENJI  
APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 724  
; LENGTH: 2443  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-094-749-724

Query Match 12.1%; Score 467; DB 13; Length 2443;  
Best Local Similarity 64.1%; Pred. No. 1.1e-115;  
Matches 755; Conservative 2; Mismatches 402; Indels 18; Gaps 3;  
QY 70 CGGGCGCCGGGGCGCGCGCTCTAGCGCTGTGTGCTGGCTGGCGCTGGCGG 129  
Db 451 CGGGCGCGCGGGCGGATGCTCGGGTGGCGGTGTGTGTGTCTGTCTGGCA 510  
QY 130 CGGGCGCGCGCCCAATACGAACGCTACAGCTTCCGAGCTTCCACCGGACGAGCTATGC 189  
Db 511 GCGCGGGGCGAGTACGAGAAGTACAGCTTCCGGGGCTTCCCGCCGAGGACCTGATGC 570  
QY 190 CGCTCGAGTCTGGCTACCGGCAACGCGTGGCAAGTACAGCGCGGAGAGTGGCGCGAGA 249  
Db 571 CGCTGGCGCGGGGTACGGGCGCTCTGGAGCAGTACGAGGAGAGAGTGGCGCGAGA 630  
QY 250 GCSTTGGCTACCTGGAGTACAGCTCGGCTGCGCTGACCGCTTGTGCGGACAGCGAGGCT 309  
Db 631 GCGCGCGCTACCTGGAGGCGGCTGCGGTGACCGCTTCTTCCGCGGACAGCGAGGCT 690  
QY 310 TCTGCCACCGCAACTGCAGCG-----CCGCGCGCGAGCGCGCGCGCGCGCG 360  
Db 691 TCTGCCACCGCAACTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750  
QY 361 TCGCAGCTATC-----CCGAGTGGCGCTCTTCCGGGGCTTGTGCGCGCGCGCGCT 414  
Db 751 GCGCAGCAGTGGCGCTTGGCGCTTCTTCCGGCGCGCTTCCAGGTGCGCTACCGCGCGCGCT 810  
QY 415 GCCTCAAGCGCTGCAAGCAGGCGCTGCGAGCTTCCGCGCGAGTCCAGCGCGCGCGGAGG 474  
Db 811 GCGTGGCGCTGCAAGCGAGCGCTGCGCGCGCTTCCAGGTGCGCTACCGCGCGCGCGAGC 870  
QY 475 TGTGGCGGAGTTCAGCGCGCGGAGCGCTTCAAGTTCCTGTGAGTTCGCTTACTTCAAGG 934  
Db 871 TGTGGCGTGAATTCAGAGCGCGCTGCGCTTACAGTACCTGCACTACGCGCTGTTCAGAG 930  
QY 535 CAAATAATCTCCCAAGGCGATCGCGCTGCTCAGACCTTCTTACTGAAAGCATCTCTGATG 594  
Db 931 CTAAACCGCTGGAGAGCGCGTGGCGCGCGCTTACACCTTCTTCCAGAGGAAACCGAGC 990  
QY 595 ACAGAAATGATGAAGAGGAAATGGCATATTTATTA---GAGCTGCTGTGTGCGGAGGACT 651  
Db 991 ACAGCTGACCGGCAAGTATCTCAACTACTATCAGGGGATGCTGGAGCTGCGCGCGAGGT 1050

QY 652 ACATTAAAGACCTGGAACCAAGTCATATGAAGCCTCTTTCATCCGAGACGTGCGGBCAT 711  
Db 1051 CCTCACGAGCCTAGAGGCCAGCCCTACGAGGCGCTGTCTCCCGGCTGTGAAGCTCT 1110  
QY 712 ACAACGGTGAAGACTGGAGAACATCCATCAGACATGGAGCTGGCCCTCCGACTTCT 771  
Db 1111 ARAACAGCGGGATTTCCGAGCAGCAGGAGGACATGAGCGGCCCTTGTACAGTACC 1170  
QY 772 TCAAAGCCCTTTACGAGTGTCTCGAGCCTCGAGGTTCCAGGAGATCAAGACTTCA 831  
Db 1171 TGGAGCTTTTGGCCCGCTGCGCGGCTGTGAAGGGGCCCATGAGCAGGTGACTTCA 1230  
QY 832 AGGATTTCTACTTTCCATAGCAGATCAATTAAGTTAGAGATTTCTGGAATGCAAAATACGT 891  
Db 1231 AGGACTTTCTACCGGCCCATAGCAGATCTCTTTCAGAGTCCCTGAGTGAAGTGGACT 1290  
QY 892 GTGAGAGAACTCACCCAGTTATAGAGGCTATCCGGTTGAGAAATTTGTGGCTACCA 951  
Db 1291 GTGAGGCCAATTTGACCCCAATGCGGTGCTACTTCTGTGACACAGTTCTGSGCCACCA 1350  
QY 952 TGTATCATTACTTGAGTTTGGCTTATTAAGTTGAGAGCCTGAAGATGACGCCCT 1011  
Db 1351 TGTACCACTACTGAGTTTGCCTACTATAAGTTGAATGATGTGCGCAGGCTGCCCGCA 1410  
QY 1012 GTGAGTCAGCTATCTGCTCTTTGATCAGAATGACAAGGTCATGCGAGCAAGCTGTGT 1071  
Db 1411 GCGCGCCAGCTACATGCTCTTCGACCCAGCAGCAGCTATGCGAGCAAGCTGTGT 1470  
QY 1072 ATTACAGTACACAGGACACTTGGGCGCTCTCGGATGAGACATTCAGGCCACAGCTG 1131  
Db 1471 ATTACGGTTCCACCGGCTCGCTGGGCGCTGGAAGAGAGGACTTCAGGCCCGGAGG 1530  
QY 1132 AAGCAGTTTCACTTTAATGTCACACACTCCAGAGGAGCTGTATGACTTTCCTAAGG 1191  
Db 1531 AGGCCATGCTTACACACACAGACCCCGGAGCTGCGGAGCTGCTGAGTTACCCACA 1590  
QY 1192 AAATATATATGATGATGATGAGGAGAGTGTGGA 1228  
Db 1591 TGTACTGTCAGTCAGATGATGATGAGTGGAGGA 1627

## RESULT 6

US-09-536-215/c  
; Sequence 215, Application US/09879536  
; Patent No. US2002014298A1  
; GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Derti, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/879,536  
; CURRENT FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: US 60/088,801  
; PRIOR FILING DATE: 1998-06-10  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 215  
; LENGTH: 618  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature

; LOCATION: (1)...(518)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-879-536-215

Query Match 12.0%; Score 466.4; DB 10; Length 618;  
Best Local Similarity 94.8%; Pred. No. 6.5e-116;  
Matches 506; Conservative 0; Mismatches 25; Indels 3; Gaps 3;  
QY 1076 CCAGTACCACAGGACACTTGGGGCTCTCGGATGAGCACTTCCAGCCGAGACCTGGAAC 1135  
Db 534 CNAGNTCCNNAGNCANTGGGGCTNTNGAATGAGCANTTCCAGCCGAGNCTTAAANC 475  
QY 1136 AGTTTCAGTTCTTAATGTGACCA-CACGCCAAGAGAGC-TGTATGACTTT-GCTAAGA 1192  
Db 474 AGTTTCAGTTCTTAATGTGNCACCACTCCAGANGAAGCTTGTATGACTTTGNTAAGA 415  
QY 1193 AAATATAATGGATGATGATGAGGAGAGTGTGGAATATGTGGATGACCTCTTGGAACT 1252  
Db 414 AAATATAATGGATGATGATGAGGAGAGTGTGGAATATGTGGATGACCTCTTGGAACT 355  
QY 1253 GGAGGAGACAGCTAGCTAGCCACAGCAACAAAGAGACTTCTCTTGGCGTTTCAGGAAACAC 1312  
Db 354 GGAGGAGACAGCTAGCTAGCCACAGCAACAAAGAGACTTCTCTTGGCGTTTCAGGAAACAC 295  
QY 1313 AGATTCCTTTGCTCTTTTCCCAACAGCCAGGCTGTTGATACCTCAGAGCCTTCTCTTAC 1372  
Db 294 AGATTCCTTTGCTCTTTTCCCAACAGCCAGGCTGTTGATACCTCAGAGCCTTCTCTTAC 235  
QY 1373 TCTCCAAAGTGAAGGAAAGCCCGCTCTCTTAACTGCATGTCATCAGGGGTGAGCCTG 1432  
Db 234 TCTCCAAAGTGAAGGAAAGCCCGCTCTCTTAACTGCATGTCATCAGGGGTGAGCCTG 175  
QY 1433 CTTTTCCTATCTCAGACCTGCACCTCATGTTCACACCTATCTTCTCACCTTTTCTT 1492  
Db 174 CTTTTCCTATCTCAGACCTGCACCTCATGTTCACACCTATCTTCTCACCTTTTCTT 115  
QY 1493 GAGATGAGTCTCGCTCTCTTCCAGGCTGGAGTGCAATGGCAGCTTCTCAGCTCAGCTG 1552  
Db 114 GAGATGAGTCTCGCTCTCTTCCAGGCTGGAGTGCAATGGCAGCTTCTCAGCTCAGCTG 55  
QY 1553 CAACCTCCGCTCTTGGGTTCAAGCAATTCCTGCTGATCAGCTCCCGAGTACC 1606  
Db 54 CAACCTCCGCTCTTGGGTTCAAGCAATTCCTGCTGATCAGCTCCCGAGTACC 1

## RESULT 7

US-09-862-832-113  
; Sequence 113, Application US/09962832  
; Patent No. US20020110821A1  
; GENERAL INFORMATION:

; APPLICANT: Ebner, Reinhard  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu  
; FILE REFERENCE: 689290-74  
; CURRENT APPLICATION NUMBER: US/09/962,832  
; CURRENT FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,077  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,280  
; PRIOR FILING DATE: 2000-09-25  
; NUMBER OF SEQ ID NOS: 259  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 113  
; LENGTH: 2347  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-862-832-113

Query Match 12.0%; Score 465.8; DB 10; Length 2347;  
Best Local Similarity 64.2%; Pred. No. 2.2e-115;  
Matches 752; Conservative 2; Mismatches 399; Indels 18; Gaps 3;  
QY 76 GCGGGGGCGCGGCGCTAGCGTGTCTGCTGCGCTGCGCTGCGCTGCGCGGCGG 135



Db 601 CGGACCTAGAGCCAGCCCTACGAGCGGTGTTCTCGGCTGTGAAGCTCTACACA 660  
Qy 718 GTGAGAACTGGAACATCCATCAGACATGAGCTGGCCCTTCCGACTTCTTCAAG 777  
Db 661 GCGGGATTTCCGAGACACCGAGGACATGAGCGGCTTGTCAAGTACCTGGCAG 720  
Qy 778 CTTTATAGAGTGTCTCGAGCTCGAGGCTTCCAGGAGATCAAGACTTCAAGGATT 837  
Db 721 TCTTTCGCGGTGCTTGGCGGCTGTGAAGGGCCCATGACAGGTGACTTCAAGACT 780  
Qy 838 TCTACCTTCCATAGACATATTAATGATGAGTCTGGAATGCTGAATGCAAAATACAGTGTGAAG 897  
Db 781 TCTACCGGCGCATAGACATCTCTTTCAGAGTCCCTCAGTGAAGTGGACTGTGAGG 840  
Qy 898 AGAAGCTCACCCAGTTATAGAGGCTATCCGGTGTGAGAAATTTGTGCTACCATGTATC 957  
Db 841 CCAATTTGACCCCAATGCGGTGCTACTTCGTGGACAAGTTCGTGGCCACCATGTACC 900  
Qy 958 ATTACTGCAAGTTTGCCTATTATAGTTGAACGACCTGAAGAATGACGCCCTGTGCAG 1017  
Db 901 ACTACTGCAAGTTTGCCTACTATAGTTGAATGATGTGCGCAGGCTGCCCGCAGCGCG 960  
Qy 1018 TCAGTATCTGCTCTTTGATCAGATGACAAGTCTATCGGTTGAGAAATTTGTGCTACCATGTATC 1077  
Db 961 CCAGTATCATGCTCTTTCGACCCCAAGGACAGGCTCATGACAGAACTGTGTATTACC 1020  
Qy 1078 AGTACCAAGGACACTTTGGGCTCTTCGGATGAGCACTTCCAGCCAGACCTTGAAGCAG 1137  
Db 1021 GGTTCACCGGCTCGCTGGGCTTGGAGAGGAGGACTTCCAGCCCGGAGGAGCCA 1080  
Qy 1138 TTCAGTTCTTAATGTGACCACTCCAGAGGAGCTGTATGACTTTGCTAAGAAATA 1197  
Db 1081 TGCTTACCAACAGACCCCGAGCTGCGGAGCTGCTGGAGTTACCCACATGTACC 1140  
Qy 1198 TAATGGATGATGAGGAGGAAGTTGTGGA 1228  
Db 1141 TGCAGTCAGATGATGAGATGAGCTGGAGGA 1171

## RESULT 9

US-09-918-995-22977  
; Sequence 22977, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 22977  
; LENGTH: 494  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-22977

Query Match 11.0%; Score 427.2; DB 11; Length 494;  
Best Local Similarity 95.0%; Pred. No. 2.6e-105;  
Matches 441; Conservative 0; Mismatches 23; Indels 0; Gaps 0;  
Qy 3259 GGGGGTCTCTCAGAGAACCTGTGAAGCCCTTATCAGTCATTATTGGCTGTGAGAA 3318  
Db 31 GGGGGTCTCTCAGAGAACCTGTGAAGCCCTTATCAGTCATTATTGGCTGTGAGAA 90  
Qy 3319 GTTCTCTGGAGTGTGGGTACATTTGAAGGCAAGTACTTCAGTTGAGGGCAAGTCTCTG 3378  
Db 91 GTTCTCTGGAGTGTGGGTACATTTGAAGGCAAGTACTTCAGTTGAGGGCAAGTCTCTG 150

Qy 3379 GAAAGAGGCTGTAGCACTTGGCAGCTACCATGCGTGTGTGTCTTGGGGTGGGGT 3438  
Db 151 GAAAGAGGCTGTAGCACTTGGCAGCTACCATGCGTGTGTGTCTTGGGGTGGGGT 210  
Qy 3439 CTGGGCACTGGCTGTGTGTGAAGGATCTTGGCAGGACACACAGCGCCCTTACTGAACCA 3498  
Db 211 CTGGGCACTGGCTGTGTGTGAAGGATCTTGGCAGGACACACAGCGCCCTTACTGAACCA 270  
Qy 3499 TCAGCATGTCAAGTGGCATTTAAGCCATGCACTGTGAGGGCCACTGAGATTGTCTCTGA 3558  
Db 271 TCAGCATGTCAAGTGGCATTTAAGCCATGCACTGTGAGGGCCACTGAGATTGTCTCTGA 330  
Qy 3559 GTATTACTGAGAACCAACAGAGCCATGATGGAGCCCTTGGGCTCTCTGGGAAAT 3618  
Db 331 GTATTACTGAGAACCATCAGATAGAGCCATGATGGAGCCCTTGGGCTCTCTGGGAAAT 390  
Qy 3619 GGGAAATCAGCCATAGGACTGAGAGGAGTACCTTAAGTTCAGAGAAACCAAGAGT 3678  
Db 391 GGGAAATCAGCCATAGGACTGAGAGGAGTACCTTAAGTTCAGAGAAACCAAGAGT 450  
Qy 3679 GTGGTGTCTTGGAGGCTGAGCTTCTTTATTCAACCTCATTTCCC 3722  
Db 451 GTGGCGGCTGGATCTGAGCTTCTTATATTAACCTCATTTCTC 494

## RESULT 10

US-09-880-107-1453/c  
; Sequence 1453, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1453  
; LENGTH: 394  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA599526  
US-09-880-107-1453

Query Match 10.1%; Score 392.4; DB 10; Length 394;  
Best Local Similarity 99.7%; Pred. No. 6.4e-96;  
Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1709 GATGATCCATCTGCTTGGCTCCACAGTCTGGGATACAGGCGTGAGCCACCATGCC 1768  
Db 394 GATGATCCATCTGCTTGGCTCCACAGTCTGGGATACAGGCGTGAGCCACCATGCC 335  
Qy 1769 CGGCTCTTCTCACCCTTTACACCTGTCTTCTTATTCCTACATCTGTTTTCACACCTTCA 1828  
Db 334 CGGCTCTTCTCACCCTTTACACCTGTCTTCTTATTCCTACATCTGTTTTCACACCTTCA 275  
Qy 1829 TCCCTGTCTTCTCATGTTCACACTGTCTTCCCATGTTCCCATGCTGCTTCTTACCA 1888  
Db 274 TCCCTGTCTTCTCATGTTCACACTGTCTTCCCATGTTCCCATGCTGCTTCTTACCA 215  
Qy 1889 TTTTGGTTGAAGGCGAGTCTTCTGCTGTGTTTTTTTTTTTTTTTCCAGAAATCAGTA 1948  
Db 214 TTTTGGTTGAAGGCGAGTCTTCTGCTGTGTTTTTTTTTTTTTTTCCAGAAATCAGTA 155  
Qy 1949 TTATTTTAAATAAGAAAAACATTCCTAGAGAATGATAATTTGTGAAAAACCTCTTTGGC 2008

Db 154 TTTATTTTAAATAAGAAAACATTCCTAGAGATGATAATTTGTGAAACCTCTTTTGGC 95  
QY 2009 TTTATTTGCTTTTCCAGATTTAGTCTCTTCTTCCCTCCATCCGGGAAAGATGTTGAAGAC 2068  
Db 94 TTTATTTGCTTTTCCAGATTTTATGCTCTCTTCTCCGCATCCGGGAAAGATGTTGAAGAC 35

QY 2069 ATAGGCTAAATTTCTCCAGCTTCAACAATGGTCTT 2102  
Db 34 ATAGGCTAAATTTCTCCAGCTTCAACAATGGTCTT 1

## RESULT 11

US-09-918-995-19090  
; Sequence 19090, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 19090  
; LENGTH: 344  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(344)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-19090

Query Match 7.5%; Score 289.4; DB 11; Length 344;  
Best Local Similarity 98.0%; Pred. No. 66-68;  
Matches 293; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1699 CTTGACTTCAGATGATCCATCTGCTTGCCTCCACAGTGTGGATTACAGCGTGAG 1758  
Db 41 CTTGACTTCAGATGATCCATCTGCTTGCCTCCACAGTGTGGATTACAGCGTGAG 100  
QY 1759 CCACCATGCCCGGCTCTTCTCAGCTTTACAGTGTCTTCTTATCTCTCATCTGTTTT 1818  
Db 101 CCACCATGCCCGGCTCTTCTCAGCTTTACAGTGTCTTCTTATCTCTCATCTGTTTT 160  
QY 1819 CACACTTCATCCCTGCTCTTCTCATGTTCACACTTGCTTCCCATCTTTCATAGCTGCC 1878  
Db 161 CACACTTCATCCCTGCTCTTCTCATGTTCACACTTGCTTCCCATCTTTCATAGCTGCC 220  
QY 1879 TTTCTTACATTTTGGTTGAGGCGAGCTCTTCTGCTGCTTTTGTGTTTTCCTCCAG 1938  
Db 221 TTTCTTACATTTTGGTTGAGGCGAGCTCTTCTGCTGCTTTTGTGTTTTCCTCCAG 280  
QY 1939 AAAATCAGTATTATTTTAAATAAGAAAAACATTCCTAGAGATGATAATTTGTGAAAA 1997  
Db 281 AAAATCAGTATTATTTTAAATAAGAAAAACATTCCTAGAGATGATAATTTGTGAAAA 339

## RESULT 12

US-09-954-456-1305/c  
; Sequence 1305, Application US/0995456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1305  
; LENGTH: 274  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-456-1305

Query Match 6.8%; Score 263.4; DB 10; Length 274;  
Best Local Similarity 97.8%; Pred. No. 61e-61;  
Matches 267; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 3599 CTTGGGCTCTCTGGGAAATGGGAAATCGAAAGGACTGAGAGGAGTTACCTTAAGG 3658  
Db 274 CTTGGGCTCTCTGGGAAATGGGAAATCGAAAGGACTGAGAGGAGTTACCTTAAGG 215  
QY 3659 TCAGAGAAAACCAAGAGAGTGTGTCTTCTGGAAGCTGAGCTTCTTTATTCAACTCAT 3718  
Db 214 TCAGAGAAAACCAAGAGAGTGTGTCTTCTGGAAGCTGAGCTTCTTTATTCAACTCAT 155  
QY 3719 TCCCTTCTCCAAATAGCCACTTGTAGTTGGGCCCTTCCAGGTTGAGGCAAGAGGA 3778  
Db 154 TCCCTTCTCCAAATAGCCACTTGTAGTTGGGCCCTTCCAGGTTGAGGCAAGAGGA 95  
QY 3779 GAAGGACAGAGCTTTGGGAAACAGAGCTTTTCTGCAATAGCTTGGGAGGAATAAAG 3838  
Db 94 GAAGGACAGAGCTTTGGGAAACAGAGCTTTTCTGCAATAGCTTGGGAGGAATAAAG 35  
QY 3839 GATAGAGTGTAAAAATAAAAAAATAAAAAA 3871  
Db 34 GATAGAGTGTGGGTTTAAAAAATAAAAAA 2

## RESULT 13

US-09-815-343-683/c  
; Sequence 683, Application US/09915343  
; Patent No. US20010055596A1  
; GENERAL INFORMATION:  
; APPLICANT: Meagher, Madeleine  
; APPLICANT: Xu, Jiangchun E.  
; APPLICANT: King, Gordon E.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
; FILE REFERENCE: 210121.504  
; CURRENT APPLICATION NUMBER: US/09/815,343  
; CURRENT FILING DATE: 2001-03-22  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 683  
; LENGTH: 279  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-815-343-683

Query Match 6.7%; Score 258.6; DB 9; Length 279;  
Best Local Similarity 97.8%; Pred. No. 1.2e-59;  
Matches 273; Conservative 0; Mismatches 4; Indels 2; Gaps 1;  
QY 1681 CCAGGCTGGTCTCGAACTTGGACTTCAGATGATCCATCTGCGCTCCACAGTGC 1740  
Db 279 CCAGGCTGGTCTCGAACTTGGACTTCAGATGATCCATCTGCGCTCCACAGAGC 220  
QY 1741 TGGATTACAGGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1800  
Db 219 TGGATTACAGGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 160  
QY 1801 TATCCTCACATCTGTTTTCACACCTT--CATCCCTGCTTCCCTCATGTTTACACATGCT 1858  
Db 159 TATCCTCACATCTGTTTTCACACCTTTCATCCCTGCTTCCCTCATGTTTACACATGCT 100  
QY 1859 TCCCATGTTTCATAGTGCCTTTCTTACCATTTTGGTTTGAAGGCGAGTCTTCTTGCT 1918  
Db 99 TCCCATGTTTCATAGTGCCTTTCTTACCATTTTGGTTTGAAGGCGAGTCTTCTTGCT 40  
QY 1919 TGTCTTTTCTGTTTTCAGAGAAATCAGTATTATTTT 1957  
Db 39 TGTCTTTTCTGTTTTCAGAGAAATCAGTATTATTTT 1

RESULT 14  
US-09-764-891-5534/c  
; Sequence 5534, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5534  
; LENGTH: 6670  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-5534

Query Match 6.2%; Score 239.2; DB 11; Length 6670;  
Best Local Similarity 86.0%; Pred. No. 1.8e-53;  
Matches 265; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 1467 ACACCTATCTTCTCACCTTTTTTTTGGATGGAGTCTCGCTCTCTTCCAGGCTGGAG 1526  
Db 4664 AGACTTTTCCAGCTTTTTTTTTTTTGAACGGAGTCTCACTCTGTCAACCCAGGCTGGAG 4605  
QY 1527 TGCAATGGCAGCTTCTCAGCTCACTGCAACTCCGCTCTTGGTTTCAAGCAATTTCTGCT 1586  
Db 4604 TACAGTGGCAGATCTCAGCTCACTGCAACTCCGCTCTTGGTTTCAAGCAATTTCTCC 4545  
QY 1587 GCATCAGCTTCCGAGTACTTGGATTACAGGATGTCACACCGCCGGCTAATTTTG 1646  
Db 4544 GCCTCAGCTTCCGAGTACTTGGATTACAGGATGTCACACCGCCAGCTAATTTTG 4485  
QY 1647 TATTTTATAGACAGCGGTTTTTGGCATGTGCGCAGGCTGGTCTCGACTCTTGACTT 1706  
Db 4484 TATTTTATAGACAGCGGTTTTTCTCCATGTGGTCTAGGCTGGTCTCGAACTCCCGACCT 4425  
QY 1707 CAGATGATCCATCTGCTTGGCTTCCACAGTGTGGGATTACAGGCGTGAGCCACCATG 1766  
Db 4424 CAGGTGATCCCGCGCTTCCGCTTCCAAAGTGTGGATTACAGGCGTGAGCCACCATGCG 4365  
QY 1767 CCGGCGCT 1774  
Db 4364 CCGGCGCT 4357

RESULT 15  
US-09-764-891-5533/c  
; Sequence 5533, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5533  
; LENGTH: 7713  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-5533

Query Match 6.2%; Score 239.2; DB 11; Length 7713;  
Best Local Similarity 86.0%; Pred. No. 2e-53;  
Matches 265; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 1467 ACACCTATCTTCTCACCTTTTTTTTGGATGGAGTCTCGCTCTCTTCCAGGCTGGAG 1526  
Db 5708 AGACTTTTCCAGCTTTTTTTTTTTTGAACGGAGTCTCACTCTGTCAACCCAGGCTGGAG 5649  
QY 1527 TGCAATGGCAGCTTCTCAGCTCACTGCAACTCCGCTCTTGGTTTCAAGCAATTTCTGCT 1586  
Db 5648 TACAGTGGCAGATCTCAGCTCACTGCAACTCCGCTCTTGGTTTCAAGCAATTTCTCC 5589  
QY 1587 GCATCAGCTTCCGAGTACTTGGATTACAGGATGTCACACCGCCGGCTAATTTTG 1646  
Db 5588 GCCTCAGCTTCCGAGTACTTGGATTACAGGATGTCACACCGCCAGCTAATTTTG 5529  
QY 1647 TATTTTATAGACAGCGGTTTTTGGCATGTGCGCAGGCTGGTCTCGAACTCTTGACTT 1706  
Db 5528 TATTTTATAGACAGCGGTTTTTCTCCATGTGGTCTAGGCTGGTCTCGAACTCCCGACCT 5469  
QY 1707 CAGATGATCCATCTGCTTGGCTTCCACAGTGTGGGATTACAGGCGTGAGCCACCATG 1766  
Db 5468 CAGGTGATCCCGCGCTTCCGCTTCCAAAGTGTGGATTACAGGCGTGAGCCACCTCGC 5409  
QY 1767 CCGGCGCT 1774  
Db 5408 CCGGCGCT 5401

Search completed: January 30, 2004, 19:07:27  
Job time : 1311.96 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 30, 2004, 05:43:24 ; Search time 471 Seconds

(without alignments)

3102.903 Million cell updates/sec

Title: US-09-729-674-2

Perfect score: 2130

Sequence: 1 MEPCRRGAALLLCVACA.....DDEGEVVEYDDLELEETS 401

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USFO\_spool/US09729674/runat\_29012004\_095315\_12934/app\_query.fasta\_1.583  
-DB=Published Applications NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09729674 @CGN 1 1 387 @runat\_29012004\_095315\_12934  
-NCPU=6 -ICPU=3 -NO.MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*  
6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
14: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
16: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
17: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
18: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
19: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
20: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

1	2128	99.9	3871	9	US-09-729-674-1	Sequence 1, Appli
2	1827.5	85.8	1810	12	US-10-108-260A-2058	Sequence 2058, Ap
3	1089.5	51.2	2443	13	US-10-094-749-724	Sequence 724, App
4	1088.5	51.1	2347	10	US-09-962-832-113	Sequence 113, App
5	1088.5	51.1	2347	12	US-10-439-388-64	Sequence 64, Appli
6	567	26.6	2600	14	US-10-045-815-3	Sequence 3, Appli
7	539	25.3	2829	14	US-10-045-815-1	Sequence 1, Appli
8	538	25.3	2322	14	US-10-045-815-7	Sequence 7, Appli
9	537	24.7	2416	14	US-10-045-815-5	Sequence 5, Appli
10	527	24.7	2753	10	US-09-728-952-28	Sequence 28, Appli
11	527	24.7	2753	10	US-09-728-952-51	Sequence 51, Appli
12	498	23.4	3396	13	US-10-034-749-921	Sequence 921, App
13	486	21.4	2865	14	US-10-071-786-86	Sequence 86, Appli
14	388	18.2	398	12	US-10-094-886-23	Sequence 23, Appli
15	321.5	15.1	2305	12	US-10-242-535A-8221	Sequence 8221, Ap
16	270	12.7	564	13	US-10-027-632-114858	Sequence 114858,
17	270	12.7	564	13	US-10-027-632-114858	Sequence 114858,
18	265.5	12.5	892	13	US-10-027-632-161474	Sequence 161474,
19	265.5	12.5	892	13	US-10-027-632-161474	Sequence 161474,
20	265.5	12.5	892	13	US-10-027-632-161476	Sequence 161476,
21	265.5	12.5	892	14	US-10-027-632-161474	Sequence 161474,
22	265.5	12.5	892	14	US-10-027-632-161475	Sequence 161475,
23	265.5	12.5	892	14	US-10-027-632-161476	Sequence 161476,
24	243	11.4	226	12	US-10-242-535A-9365	Sequence 9365, Ap
25	200.5	9.4	349	9	US-09-925-302-395	Sequence 395, App
26	195	9.2	618	10	US-09-879-536-215	Sequence 215, App
27	168	7.9	150	12	US-10-242-535A-1775	Sequence 1775, Ap
28	161.5	7.6	549	13	US-10-029-386-2358	Sequence 2358, Ap
29	156	7.3	181	13	US-10-029-386-16058	Sequence 16058, A
30	121	5.7	247	12	US-10-242-535A-4919	Sequence 4919, Ap
31	121	5.7	1584	10	US-09-974-300-2416	Sequence 2416, Ap
32	119	5.6	342	12	US-10-242-535A-12155	Sequence 12155, A
33	105	4.9	1239	12	US-10-359-493-47173	Sequence 47173, A
34	101	4.7	1230025	12	US-10-289-782-1	Sequence 1, Appli
35	100	4.7	5457	13	US-10-161-051-98	Sequence 98, Appli
36	99.5	4.7	3899	13	US-10-360-522-53	Sequence 53, Appli
37	99	4.6	2238	13	US-10-032-585-6860	Sequence 6860, Ap
38	99	4.6	4643	10	US-09-864-864-266	Sequence 266, App
39	98.5	4.6	2301	12	US-10-104-047-1504	Sequence 1504, Ap
40	98	4.6	65	13	US-09-908-975-28555	Sequence 28555, A
41	97.5	4.6	2285	15	US-10-198-846-9848	Sequence 9848, Ap
42	96.5	4.5	2249	11	US-09-946-374-117	Sequence 117, App
43	96.5	4.5	2249	12	US-10-011-795A-117	Sequence 117, App
44	96.5	4.5	2249	12	US-10-012-231A-117	Sequence 117, App
45	96.5	4.5	2249	13	US-10-015-387A-117	Sequence 117, App

#### ALIGNMENTS

#### RESULT 1

US-09-729-674-1  
Sequence 1, Application US/09729674

Patent No. US2001003935A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John M.

APPLICANT: LaVallie, Edward R.

APPLICANT: Collins-Racie, Lisa A.

APPLICANT: Evans, Cheryl

APPLICANT: Merberg, David

APPLICANT: Treacy, Maurice

APPLICANT: Agostino, Michael J.

APPLICANT: Steining, Robert J.

APPLICANT: Spaulding, Vikki

APPLICANT: Wong, Gordon G.

APPLICANT: Clark, Hilary

APPLICANT: Pechtel, Kim

APPLICANT: Genetics Institute, Inc.

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

FILE REFERENCE: 6055-64X

CURRENT APPLICATION NUMBER: US/09/729,674

CURRENT FILING DATE: 2000-12-04

PRIOR APPLICATION NUMBER: 09/539,330















```
Db 234 TCGCGCACGCTGTGCCACCGAAGTCCGCGTGGCCAGCGACCTGGATCTCGTCCGGAC 293
Qy 96 ProAla-----AlaGlyLeuAlaSerTyrProGluLeuArgLeuPheGlyGly 111
Db 294 CCCAGCTGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 353
Qy 112 LeuLeuArgAlaHisCysLeuLysArgCysLysGlnGlyLeuProAlaPheArgGln 131
Db 354 GTGCTGCGCGTGGCGCGCTGCTAGCGCGTGC---CTCGGCGCGCGCTCTGCCACTG 410
Qy 132 SerGlnProSerArgGluValLeuAlaAspPheGlnArgGluProTyrLysPheLeu 151
Db 411 CTG-----AGTGAAGAACTGACCTGGAGTTCAACAGCGGCGCGCGTACAACTAC 464
Qy 152 GlnPheAlaTyrPheLysAlaHisCysLeuLysArgCysLysGlnGlyLeuProAlaPhe 171
Db 465 CAGGTGCGCTATTTCAGATAAACAAGCTGGAGAAAGCTGTGCTGCGCGCACACCTTC 524
Qy 172 LeuLeuLysHisProAspAspGluMetLysArgAsnMetAlaTyrTyrLysSerLeu 191
Db 525 TTGTGGCAATCCGAGACATGAGATGCGGAGAACCTCGACTATTACCAACCATG 584
Qy 192 ProGlyAlaGluAspTyr---ileLysAspLeuGluThrLysSerTyrGluSerLeuPhe 210
Db 585 TCTGGGTGAGGAGCAGACTTCAGGATCTCGAGGCCAAGCCCATATGCATGAGTTT 644
Qy 211 IleArgAlaValArgAlaTyrAsnGlyGluAsnTyrPArgThrSerIleThrAspMetGlu 230
Db 645 CGGCTGGGGTACGACTCTTACTCAGAGAGAACCCACAGGAGCTGTGCCACCTGGAG 704
Qy 231 LeuAlaLeuProAspPhePheLysAlaPheTyrGluCysLeuAlaAlaCysGluGlySer 250
Db 705 GCGGCACTGACAGAGTACTTTTGCCCATGAGAGTGCCTGCTCGAGAGGGGCC 764
Qy 251 ArgGluLeuLysAspPheLys-----AspPheTyrLeuSerIleAla 264
Db 765 TATGACTACGCGGTACAACTACCTAGACTACAGCGCTGACCTCTCCAGGCCATCACA 824
Qy 265 AspHisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsnLeu---ThrPro 283
Db 825 GATCATTTAGCTCAGGTCTCACTGTAAGCAAGAACTGTGTCACGAGCTGGCTTCCAC 884
Qy 284 ValIleGlyGlyTyrProValGluLysPheValAlaThrMetTyrHisTyrLeuGlnPhe 303
Db 885 CCAAGTAGGGAAGAAAGCCCTTTGAAGACTTCCTCCCTTCACACTATATTACTACAGTTT 944
Qy 304 AlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAlaProCysAlaValSerTyrLeuLeu 323
Db 945 GCCTACTACAACATTGGGAAGCTATACACAGCTATTGAATGTGCCAAGACCTACTCTCTC 1004
Qy 324 PheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyrGlnTyrHisArgAsp 343
Db 1005 TTCTTTCCATGATGAGGTGATGACACAGAACTGTTGGCTTATTAC----- 1049
Qy 344 ThrTrpGlyLeuSerAspGluHis-----PheGlnProArgProGluAlaValGln 360
Db 1050 ACAGCCATCTCTGGAGAAGAGAGAGCCAGCTCCATCAGCCCGCAGGAGATGCGGAGAA 1109
Qy 361 PhePheAsnValThrThrLeuGlnLysGluLeu-----TyrAspPheAlaLys 376
Db 1110 TACCAGCTGCCAAACCTGTTGGAGAAAGAACTGCTTTTCTCGCTTATGACATTTTGGGA 1169
Qy 377 GluAsnIleMetAspAspAspGluGluValValGluTyrValAspAspLeuGlu 396
Db 1170 ATTCCCTTTGTGATCCCGATTCATGGACTCCAGAGAAGTGAATCCCAAGAGATTGCAA 1229
Qy 397 LeuGluGluThrSer 401
Db 1230 GAGAAACAGAGACTCT 1244
```

RESULT 9

US-10-045-815-5

```
; Sequence 5, Application US/10045815
; Publication No. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadawa, Renu
; APPLICANT: Sugihara, Takashi
; APPLICANT: Oribe, Akiko
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCES: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045,815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JP00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118906
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(2252)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2376..feature
; OTHER INFORMATION: n = A,T,C or G
; US-10-045-815-5
Alignment Scores:
Pred. No.: 1,38e-57 Length: 2416
Score: 538.00 Matches: 147
Percent Similarity: 48.71% Conservative: 60
Best Local Similarity: 34.59% Mismatches: 173
Query Match: 25.26% Indels: 46
DB: 14 Gaps: 13
US-09-729-674-2 (1-401) x US-10-045-815-5 (1-2416)
```

```
Qy 5 ArgArgGlyAlaAlaAlaLeuAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 24
Db 22 CGAAGAGAGAGCTGCTGGCAGATGCTAGCGGTGCGCGCGCGCGCGCGCGCGCGCG 81
Qy 25 ArgAlaGlnTyrGluArgTyrSerPheArgSerPheProArgAsp---GluLeuMetPro 43
Db 82 -----GCGAGTCTGAGCGCGGATGGGACGTGGCAGCC 114
Qy 44 LeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGlyGluHisTrpAlaGluSer 63
Db 115 CTGACCTG-CTTTACGACAGAGGAGCGCGGCTACTCGCGCAGGAGACTGGCCCGGGTG 173
Qy 64 **GlyTyrLeuGluIleSerLeuArgLeuHisArgLeuLeuArgAspSerGluAlaPhe 83
Db 174 GTCTGAACATGAGAGCGGCTGTGCTCGCGGGGGCGCTGCTGCTGCTGCTGCTGCTGCTG 233
Qy 84 CysHisArgAsnCysSerAla-----AlaProGln-----ProGlu 95
Db 234 TGCGGCACACGCTGTGCCACCGAAGTCCGTCGGCGCACCGGACCTGGATCTCGTCCGGAC 293
Qy 96 ProAla-----AlaGlyLeuAlaSerTyrProGluLeuArgLeuPheGlyGly 111
Db 294 CCCAGCTGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 353
Qy 112 LeuLeuArgAlaHisCysLeuLysArgCysLysGlnGlyLeuProAlaPheArgGln 131
Db 354 GTGCTGCGCGTGGCGCGCTGCTAGCGCGTGC---CTCGGCGCGCGCTCTGCCACTG 410
Qy 132 SerGlnProSerArgGluValLeuAlaAspPheGlnArgGluProTyrLysPheLeu 151
Db 411 CTG-----AGTGAAGAACTGACCTGGAGTTCAACAGCGGAGCGCGCTACAACTAC 464
Qy 152 GlnPheAlaTyrPheLysAlaHisCysLeuLysArgCysLysGlnGlyLeuProAlaPhe 171
```



RESULT 10

```

US-09-728-952-28
; Sequence 28, Application US/09728952
; Patent No. US20020111302A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Yamazaki, Vicki
; APPLICANT: Ujwal, Manusha L.
; APPLICANT: Drmanac, Radcoje T.
; TITLE OF INVENTION: No. US20020111302A1 Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 799
; CURRENT APPLICATION NUMBER: US/09/728,952

```

```
QY 186 AlaTyrTyrLysSerLeuProGluAlaGluAspTyr---IleLysAspLeuGluThrLys 204
Db 760 GACTATTACCAACCATCTCTGGAGTGAAGGCGCGACTTCAAGGATCTTGAGACTCAA 819
QY 205 SerTyrGluSerLeuPheIleAlaValAlaTyrAsnGlyGluAsnTyrArgThr 224
Db 820 CCCCATATGCAAGAAATTTGCACTGGAGTGGGACTTACTCAGAGGAACACCCACAGAA 879
QY 225 SerIleThrAspMetGluLeuAlaLeuProAspPheLysAlaPheTyrGluCysLeu 244
Db 880 GCTGTGCCCCACCTAGAGCGCGCTGCAAGAACTTTTGGCTATGAGGAGTGGCGT 939
QY 245 AlaAlaCysGluGlySerArgGluLeuLysAspPheLys-----Asp 258
Db 940 GCCCTCTGCGAAGGCGCCCTATGACTACGATGGCTACAACTACCTTGGTACACGCTGAC 999
QY 259 PheTyrLeuSerIleAlaAspHisTyrValGluValLeuGluCysIleGluCysGlu 278
Db 1000 CTCCTCCAGGCGCATCAGAGATTTATCATCCAGGCTCTCACTGTAAAGCAAACTGTGTC 1059
QY 279 GluAsnLeu---ThrProValIleGlyTyrProValGluLysPheValAlaThrMet 297
Db 1060 ACGGAGCTTCTTCCACCAAGTCGAGAGAGCCCTTTGAAGACTTCTCCCTCCCATGCAT 1119
QY 298 TyrHisTyrLeuGlnPheAlaTyrTyrLysLeuAsnAspLeuLysAsnAlaProCys 317
Db 1120 TATAATTATCTGCACTTTGCCCTACTATACATTTGGGAATTATACACAGGCTGTTGAATGT 1179
QY 318 AlaValSerTyrLeuLeuPheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyr 337
Db 1180 GCCAAGACCTATCTTCTCTTCTCCCAATGACGAGGATGATGAACCAAAATTTGGCCTAT 1239
QY 338 TyrGlnTyrHisArgAspThrTyrGlyLeuSerAspGluHisPheGln-----Pro 354
Db 1240 TATGAGCTATG-----CTTGGAGGAAGAACACACAGATCCATCGGCGCC 1284
QY 355 ArgProGluAlaValGlnPhePheAsnValThrLeuGlnLysGluLeuTyrAspPhe 374
Db 1295 CGTGAGAGTGCCAAAGAGTACCGACGAGGAGCTACTTGGAAAGAACTGCTTTCTTC 1344
QY 375 AlaLysGlu-----AsnIleMetAspAspGluGlyGluValValGluTyr 390
Db 1345 GCTTATGATGTTTGTGAATTCCTTCTGTTGGATCCGATTCATGAGCTCCAGAGAAGTG 1404
QY 391 ValAspAspLeuLeuGluLeuGluThrSer 401
Db 1405 ATTCACAGAGATTGCAAGAGAACAGAGTCA 1437

RESULT 11
US-09-728-952-51
; Sequence 51, Application US/09728952
; Patent No. US2002011302A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Kyle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Yamazaki, Vicki
; APPLICANT: Ujwal, Manusha L.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US2002011302A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/728,952
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 51
; LENGTH: 2753
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2418)
; US-09-728-952-51

Alignment Scores:
Pred. No.: 4,35e-56 Length: 2753
Score: 527.00 Matches: 145
Percent Similarity: 43.38% Conservative: 68
Best local Similarity: 29.53% Mismatches: 188
Query Match: 24.74% Indels: 110
DB: 10 Gaps: 12

US-09-729-674-2 (1-401) x US-09-728-952-51 (1-2753)

QY 6 ArgGlyValAlaAlaLeuLeuAlaLeuLeuCysValAlaCysAlaLeuArgAlaGlyArg 25
Db 10 CGCGCGCTTGAAGCTGCTGACCACTGCTGGGTGCTGGCGCGCT-----GCCTCCCAA 63
QY 26 AlaGlnTyrGluArgTyrSerPheArgSerPheProA:GAspGluLeuMetProLeuGlu 45
Db 64 GCCGAGTCCGAGTCCGAGGACGATGGGCGCATGGTGAAGCGCTGATCTGCTC----- 114
QY 46 SerAlaTyrArgHisAlaLeuAspLysTyrSerGlyGluHisTyrAlaGluSer**Gly 65
Db 115 -----TTCCGCGAGGGGACCCGACCTACGCGCGCGGGGACTGGCCGGGTGCTCTG 168
QY 66 TyrLeuGluIleSerLeuArgLeuHisArgLeuLeuArgAspSerGluAlaPheCysHis 85
Db 169 AGCATGAACGGGCGCTGGCTCCCGGGCAGCCCTCCCGCGCGCTTCGCGCTGCTGCGCGC 228
QY 86 ArgAsnCysSerAla-----AlaProGlnProGlu 95
Db 229 ACCAGCTGTGCGCGCGACTTCCCGTGGAGCTGACCCGCGACTGTGCTCCCGCGCGCC 288
QY 96 ProAlaAlaGlyLeuAlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArg 115
Db 289 CAGGCTCTGGGCGCGCGCGCTGCGGACCTGAGCTTCTCGGGGCGCTTCGCGGTGCG 348
QY 116 AlaHisCysLeuLysArgCysLysGlnGlyLeuProAlaPheArgGlnSerGlnProSer 135
Db 349 GCTGCTGCTGCGCGCTGCTC---CTCGGCGCGCGCGCGCGCTCGCTC-----AGC 399
QY 136 ArgGluValLeuAlaAspPheGlnArgGluProTyrLysPheLeuGlnPheAlaTyr 155
Db 400 GAAGATGAGCTGGAGTTCCGCAAGCGGAGGCCCTTACACTACTGCGAGTCCGCTAC 459
QY 156 PheLysAla----- 158
Db 460 TTCAAGTGCAGACCTGCTCGAAACCGAGCGCGCGGGTCTTCTGGGAGAGAGTGT 519
QY 158 ----- 158
Db 520 GCAGGGACCTGAGGAGCTTGGGGGATCGGGAAGTGTCCGCGAGGGGAAAGTGGCC 579
QY 158 ----- 158
Db 580 TCCTGCTGGGAGCTCTCTCTCGAGCGGGGAGAGTGTCTCTCGGAGGAGACTTCC 639
QY 159 -----AsnAsnLeuProLysAlaIle 165
Db 640 TCGCCCATTCGCATGGCGAGATGCTTAACCCCAAAGATCAACAAGTGGAGAACTGT 699
QY 166 AlaAlaAlaHisThrPheLeuLeuLysHisProAspAspGluMetMetLysArgAsnMet 185
Db 700 GCTGAGCAGACACACTTCTTCTGCGCAATCTGAGCACATCGGAATGAGAGAACCTA 759
QY 186 AlaTyrTyrLysSerLeuProGlyAlaGluAspTyr---IleLysAspLeuGluThrLys 204
Db 760 GACTATTACCAACCATGCTGAGTGAAGGAGGCGGACTTCAAGGATCTTGAGACTCAA 819
```



```

Db 972 AGAATGAGTATTTAGGGTATAGCGTGTGTATGAGCTATTCAGATCTACTACAT 1031
QY 268 IGLUValLeuGluCysLysIleGlnCysGluGluAsnLeuThrProValIleGlyGlyTy 288
Db 1032 CGAGTGCTTGTTCAGCATGATGTGTGAGGAACTTCCACCCCGCCCTGCGCGCT 1091
QY 288 x---ProValGluLysPheValAlaThrMetTyrHisTyrLeuGlnPheAlaTyrTyrLy 307
Db 1092 CTCCTCCCATCGAAGATTTCTTCCTCGCACTATGATACCTACAGTTTGCTACTATCG 1151
QY 307 sLeuAsnAspLeuLysAsnAlaProCysAlaValSerTyrLeuLeuPheAspGlnAs 327
Db 1152 AGTGTGAGTATGTGAAGCCCTCGAGTGTGCCAAGCCCTATCTTCATGCCATCCAGA 1211
QY 327 nAspLysValMetGlnGlnAsnLeuValTyrGlnTyrHisArgAspThrTrpGlyLe 347
Db 1212 TGATGAGGATGTCTTGCAGCAATGTGCACTACTATGAG-----AGTCT 1253
QY 347 userAspGluHisPheGlnPro-----ArgProGluAlaValGlnPhePh 362
Db 1254 GCTGATGATAGCATGTGCCCGGCATCCATTGAGCCAGAGAGATTTAAACAATGTTGT 1313
QY 362 eAsnValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGluAsnIle 379
Db 1314 GAAAGTCAATAGCTGGAGTCTGAGCTGATAAATCAGCTGCAGAGGCTCG 1365

```

## RESULT 13

```

US-10-071-766-86
; Sequence 86, Application US/10071766
; Publication No. US20020192678A1
; GENERAL INFORMATION:
; APPLICANT: Huei-Mei Chen
; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE
; FILE REFERENCE: PA-0043 US
; CURRENT APPLICATION NUMBER: US/10/071,766
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: PERL Program
; SEQ ID NO 86
; LENGTH: 2665
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020192678A1 978730.4
; NAME/KEY: unsure
; LOCATION: 353-387, 694, 727
; OTHER INFORMATION: a, t, c, g, or other
US-10-071-766-86

```

```

Alignment Scores:
Pred. No.: 4,898-47 Length: 2665
Score: 456.00 Matches: 123
Percent Similarity: 47.26% Conservative: 58
Best Local Similarity: 32.11% Mismatches: 143
Query Match: 21.41% Indels: 59
DB: 14 Gaps: 11

```

US-09-729-674-2 (1-401) x US-10-071-766-86 (1-2665)

```

QY 19 CysAlaLeuArgAlaGlyArgAlaGlnTyrGluArgTyrSerPheArgSerPheProArg 38
Db 470 TGCTCTAGCCAGCGCGCGCGCGCGCTAC-----499
QY 39 AspGluLeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGlyGlu 58
Db 500 -----TACAGCGGAGAC 511
QY 59 HisTrpAlaGluSer***GlyTyrLeuGluLeuSerLeuArgLeuHisArgLeuLeuArg 78
Db 512 ---TACGACGAGCGCGCGCGCGCGCTGGAAGCGCGCTGCGCAGCCACCGCGCGCTCGG 568

```

```

QY 79 AspSerGluAlaPheCysHisArg-AsnCysSerAla-----AlaProG1 93
Db 569 GAAATCCGACGCGCTGTCCCTGCGCTGCGCGCGCCACCGCTCTCCCGCCCCCGC 628
QY 93 nProGluProAlaAlaGlyLeuAlaSerTyrProGluLeuArgLeuPheGlyGlyLeu 113
Db 629 CCCCCCGCGCGCGCGCGCGCTGAGTGC-----CTTTTCGCTCCTGTGT 679
QY 113 uArgArgAlaHisCysLeuLysArgCys-----LysGlnGlyLeuProAlaPheAr 130
Db 680 GGGCGCGCGCGCTNTTATCGCAGCTGTGAGACCCGCGCTCGGGGNCCTCCGATCCCG 739
QY 130 gGlnSerGlnProSerArgGluValLeuAlaAspPheGlnArgArgGluProTyrLysPh 150
Db 740 CCAC---CCGCTCAGCGAGATGTGCGCAGCACTTCCAGCGCAGAGTGCCTCAACTA 796
QY 150 eLeuGlnPheAlaTyrPheLysAlaAsnAsnLeuProLysAlaIleAlaAlaHisPh 170
Db 797 CTGCGAGCGGCTACATCAAGCTTAACAGCTCGAANAAGCAGTGGAGCGAGCTCACAC 856
QY 170 rPheLeuLeuLysHisProAspAspGluMetMetLysArgAsnMetAlaTyrTyrLysSe 190
Db 857 ATTTTCGTGGCTAACCTGAGCAGCATGGAATGCGCAGCAACATTGAGATTACAGGC 916
QY 190 rLeuProGlyAlaGluAsp---TyrIleLysAspLeuGluThrLysSerTyrGluSerIe 209
Db 917 GACAGCTGTGTGAAGCATTCGAGTTGTAGACAGAGAACCAAGCCACACATGGAGAG 976
QY 209 uPheIleArgAlaValArgAlaTyrAsnGlyAsnTrpArgThrSerIleThrAspMe 229
Db 977 TTACAATGAGGAGTTAAACATTATGAGCTCATGACTTTGAGATGGCTTATCAGGCATT 1036
QY 229 tGluLeuAlaLeuProAspPhePheLysAlaPheTyrGluCysLeuAlaAlaCysGluGl 249
Db 1037 CGAACCAAGCCTTAAGAGAAATATTCTGTTGAAGATACAGATGCCGACCCCTATGTGAGG 1096
QY 249 ySerArgGluIleLysAspPheLys-----AspPheTyrLeuSerI1 263
Db 1097 GCCTCAGAGATTGAAGATATGAGTATTAGGGTATAAGGCTGCTGTATGAGCTAT 1156
QY 263 eAlaAspHisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsnLeuThrPr 283
Db 1157 TGCAGATCACTACATGCGAGTCTTGTGTCAGCATGAATGTGTGAGGGAACCTTGCCAC 1216
QY 283 oValIleGlyGlyTyr---ProValGluLysPheValAlaThrMetTyrHisTyrLeuGl 302
Db 1217 CGCCCTGCGCGCTCTCTCCCTCGAAGATTTCTTCTCTGCACTATGATTACTACA 1276
QY 302 nPheAlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAlaProCysAlaValSerTyrLe 322
Db 1277 GTTGCCTACTATCGAGTTGTGAGTATGTGAAGCCCTCGAGTGTGCGAAGCCTATCT 1336
QY 322 uLeuPheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyrTyrGlnTyrHisAr 342
Db 1337 TCTATGCCATCCAGATGATGAGGATGCTCTAGACAAATGTTGATGATTACTATGAG----- 1389
QY 342 gAspThrTrpGlyLeuSerAspGluHisPheGlnPro-----ArgProG1 357
Db 1389 -----AGTCTGCTGATGATAGCATTTGACCCGCGCATCCATTGAGGCGCAGAGGA 1438
QY 357 uAlaValGlnPhePheAsnValThrLeuGlnLysGluLeuTyrAspPheAlaLysGl 377
Db 1439 TTAAACAATGTTTGAACAACGTCATAGCTGAGTGTGAGCTGATAAATCAGCTGCAGA 1498
QY 377 uAsnIle 379
Db 1499 AGGTCTG 1505

```

## RESULT 14

```

US-10-242-535A-8221
; Sequence 8221, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:

```

```
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 8221
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2)..(2)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (263)..(263)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-8221

Alignment Scores:
Pred. No.: 1,04e-39 Length: 398
Score: 388.00 Matches: 81
Percent Similarity: 72.57% Conservative: 1
Best Local Similarity: 71.88% Mismatches: 1
Query Match: 18.22% Indels: 30
DB: 12 Gaps: 1

US-09-729-674-2 (1-401) x US-10-242-535A-8221 (1-398)

Qy 291 GlulysPheValAlaThrMetTyrHisTyrLeuGlnPheAlaTyrTyrLysLeuAsnAsp 310
Db 61 GAGAAATTGGCTACCATGATATATTACTGTCAGTCTCTTTGATCAGATGACAGGAC 120

Qy 311 LeulysAsnAlaAlaProCysAlaValSerTyrLeuLeuPheAspGlnAsnAspLysVal 330
Db 121 CTGAAGATGACAGCCCTGTGCAGTCAGCTATCTCTTTGATCAGATGACAGGTC 180

Qy 331 MetGlnGlnAsnLeuValTyrGlnTyrHisArgAspThrTrpGlyLeuSerAspGlu 350
Db 181 ATGACAGCAGAACCTGGTGTATTACAGTACCACAGGACACGTGGGGCCCTCTCAGATGAG 240

Qy 351 HisPheGlnProArgProGluAlaValGlnPhePheAsnValThrThrLeuGlnLysGlu 370
Db 241 CACTTCAGCCCAACCT----- 258

Qy 371 LeuTyrAspPheAlaLysGluAsnIleMetAspAspGluGly-GluValValGlnTyr 390
Db 259 -----GGAGNAAGTTGTGGAATA 276

Qy 390 rVal-AspAspLeuLeuGluLeuGluGlnThrSer 401
Db 277 TGTGGATGACCTCTTGGAACTGGAGGAGACCAGC 311

RESULT 15
US-10-094-886-23
; Sequence 23, Application US/10094886
; Publication No: US20040002120A1
; GENERAL INFORMATION:
; APPLICANT: Kameda, Ramesh
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A.
```

```
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Catherine
; APPLICANT: Vernet, Corine A.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Miller, Charles
; APPLICANT: Casman, Stacie
; APPLICANT: Pena, Carol
; APPLICANT: Gangolli, Esha
; APPLICANT: Gusev, Vladimir
; APPLICANT: Smithson, Glennnda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Gerlach, Valerie
; APPLICANT: Pochart, Pascal
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimketa, Richard
; APPLICANT: Rastelli, Luca
; APPLICANT: Spaderna, Steven
; APPLICANT: LaRoche, William
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
; FILE REFERENCE: 21402-290 B
; CURRENT APPLICATION NUMBER: US/10/094,886
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,182
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/288,052
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: Patent in 2.1
; SEQ ID NO 23
; LENGTH: 2305
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (36)..(2244)
US-10-094-886-23

Alignment Scores:
Pred. No.: 5.96e-30 Length: 2305
Score: 321.50 Matches: 103
Percent Similarity: 43.67% Conservative: 59
Best Local Similarity: 27.76% Mismatches: 156
Query Match: 15.09% Indels: 53
DB: 12 Gaps: 12

US-09-729-674-2 (1-401) x US-10-094-886-23 (1-2305)

Qy 48 TyrArgHisAlaLeuAspLysTyrSerGlyHisTrpAlaGluSer**GlyTyrLeu 67
```

```
Db 1197 AGCAGCTCTAC---TATGCCATGGAGCACCTG 1226

Search completed: January 30, 2004, 08:05:27
Job time : 500 secs

156 TACGCTGACGGCTGCGCGCTACGCGCGCGGGCTTGGCGCGCGCGCTGGCGCTG 215
Qy 68 GlutLeuSerLeuArgLeuHisArgLeuLeuArgAspSerGluAlaPheCysHisArgAsn 87
Db 216 CGGAGGCGCTGCGGAGCAGCGCGCGCTGGCGCGGGTGGCGCTGGCGTGGCGGCGAGC 275
Qy 88 CysSerAlaAlaPro-----GlnProGluProAla 97
Db 276 TGGCGGCGGATCGGCGCGCGCTCCCGCGCTTCTCGGGCGCGCGGAGCCCGAC 335
Qy 98 Alagly-----LeuAlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeu 113
Db 336 TCCGGCGCGGACCCACGAGGGGTCCTGG---GAGCGACAGCTTCTCCGTGCGCGCTC 392
Qy 114 ArgArgAlaHisCysLeuArgCysLysGln-----GlyLeuPro 127
Db 393 CGCGCGCGGACTGCTCACCAGTGGCGACGAGGCTGGCGCGCGGGGCGCGCG 452
Qy 128 AlaPheArgGlnSerGlnProSerArgGluValLeuAlaAspPheGlnArgGluPro 147
Db 453 CGGCTTCGCGTGGGAGCGCGCTCGGGAGCGC-----TTCCGCGCTCGGGAGCCC 503
Qy 148 TyrLysPheLeuGlnPheAlaTyrPheLysAlaAsnAsnLeuProLysAlaAlaAla 167
Db 504 TACAACCTACCTGACAGGCGCTATTACCGATTGAAGAGCTGGATCTGGCAGCTGCGGCA 563
Qy 168 AlaHisThrPheLeuLeuLysHisProAspAspGluMetMetLysArgAsnMetAlaTyr 187
Db 564 GCACACACCTTCTTTGTAGCAAAACCCATGCACCTGCAGATCGCGGAGGACATGGCTAAG 623
Qy 188 TyrLysSerLeuProGlyAlaGlu---AspTyrIleLysAspLeuGluThrLysSerTyr 206
Db 624 TACAGCAATGTCGGGAGTTCGGCCCGCAGAGCTTCCGGGACCTGGAGACCGCCCCACAC 683
Qy 207 GluSerLeuPheIleArgAlaValArgAlaTyrAsnGlyGluAsnTrpArgThrSerIle 226
Db 684 TGGCGAGCTATGACACTGGCTGAGCTACTGGGGCGCGCAGGAGGAGGAGGAGTGGCGACTG 743
Qy 227 ThrAspMetGluLeuAlaLeuProAspPhePheLysAlaPheTyrGluCysLeuAlaAla 246
Db 744 CCCAGGCTAGAGAGGAGCTCTTCAGGGGAGCGCTGCGCCAGATGGAGAGCTGCGCTGCAC 803
Qy 247 CysGluGlySerArgGluIleLysAspPheLysAsp----- 258
Db 804 TGTGAGGGCGCTGAGAGCAGCAGCGGGGCTGAGAGAGAGGAGGATGGGGCTGCGAGCCAG 863
Qy 259 -----PheTyrLeuSerIleAlaAspHisTyrValGluValLeuGluCysLysIleGln 276
Db 864 GGGGGCGCTCTATGAGGCCATTGCAGGACACTGGATTTCAGTCTGCTGCGCGGCAAGC 923
Qy 277 Cys---GluGluAsnLeuThrProValIleGlyTyrProValGluLysPheValAla 295
Db 924 TGTGTGGGGGAGAGCAGCCGCTGCTGGTGCAGCTTCCCTGTCGCCAGACTTCTTCCC 983
Qy 296 ThrMetTyrHisTyrLeuGlnPheAlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAla 315
Db 984 AACCAAGCTGAGCGCGCTACATGAGCCCATGCTCAGTGGGCAATCTGTCCAGGCTATA 1043
Qy 316 ProCysAlaValSerTyrLeuLeuPheAspGlnAsnAspLysValMetGlnGlnAsnLeu 335
Db 1044 GAAATGTCCTGAGTGTCTGCTCTTACCCGGAGGATGAGGCTGCCAGAGGGGCTCTG 1103
Qy 336 ValTyrTyrGlnTyrHisArgAspThrTrpGlyLeuSerAspGluHisPheGlnProArg 355
Db 1104 AACCAGTACCGCCAGCTGGGA-----GAGCCGAGA 1136
Qy 356 Pro-----GluAlaValGlnPhePheAsnValThrLeu-----Gln 368
Db 1137 CCTGGCCTCGGACCCAGAGGAGACATCCAGCGCTTTCATCCTCCGATCCTCGGGGAGAG 1196
Qy 369 LysGluLeuTyrAspPheAlaLysGluAsnIle 379
Db 369 LysGluLeuTyrAspPheAlaLysGluAsnIle 379
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 30, 2004, 05:18:19 ; Search time 81 Seconds  
(without alignments)

2185.119 Million cell updates/sec

Title: US-09-729-674-2

Perfect score: 2130

Sequence: 1 MEPCRRGAALLALLVCACA.....DDEGEVVEYVDLLELETS 401

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09729674/runat.29012004.095314.12909/app.query.fasta\_1.583  
-DB=Issued Patents.NA -QPWT=fastap -SURFIX=p2n.rni -MINMATCH=0.1 -LOOPCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09729674 @cgn2\_1.1.56 @runat.29012004.095314.12909 -NCFU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELCP=6 -DELEXT=7

Database :

Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	195	9.2	618	3	US-09-328-111-215 Sequence 215, Appl
2	140	6.6	11309	4	US-08-961-527-108 Sequence 108, Appl
C 3	106	5.0	1664976	4	US-08-916-421B-1 Sequence 1, Appli
4	102.5	4.8	80161	3	US-09-036-987A-1 Sequence 1, Appli
5	102.5	4.8	80161	3	US-09-370-700-1 Sequence 1, Appli
6	102.5	4.8	80161	3	US-09-603-207-1 Sequence 1, Appli
7	101.5	4.8	1131	4	US-09-252-991A-7006 Sequence 7006, Ap
8	101	4.7	1230025	4	US-09-198-452A-1 Sequence 1, Appli
9	100.5	4.7	1557	4	US-09-252-991A-7033 Sequence 7033, Ap
C 10	100.5	4.7	2751	4	US-09-252-991A-7066 Sequence 7066, Ap
11	100.5	4.7	5833	4	US-09-221-0175-545 Sequence 545, App
12	99.5	4.7	2316	4	US-09-252-991A-6950 Sequence 6950, Ap

13	97	4.6	963	4	US-09-252-991A-6979 Sequence 6979, Ap
14	97	4.6	2233	4	US-09-252-991A-15346 Sequence 15346, A
C 15	97	4.6	2453	4	US-09-252-991A-15430 Sequence 15430, A
16	97	4.6	2784	4	US-09-252-991A-15252 Sequence 15252, A
17	96.5	4.5	2079	4	US-09-489-847-25 Sequence 25, Appl
C 18	95.5	4.5	1230	4	US-09-252-991A-10343 Sequence 10343, A
19	95.5	4.5	1389	4	US-09-252-991A-10807 Sequence 10807, A
20	94.5	4.4	4184	2	US-08-785-310A-4 Sequence 4, Appli
21	94.5	4.4	8948	4	US-09-643-597-119 Sequence 119, App
22	94.5	4.4	8948	4	US-09-480-884A-119 Sequence 119, App
23	94.5	4.4	8948	4	US-09-542-615A-119 Sequence 119, App
24	94.5	4.4	8948	4	US-09-606-421B-119 Sequence 119, App
25	94	4.4	32798	4	US-09-604-694B-1 Sequence 1, Appli
26	94	4.4	34303	2	US-08-735-609-4 Sequence 4, Appli
27	94	4.4	34303	2	US-08-735-609-4 Sequence 4, Appli
28	94	4.4	34303	3	US-09-315-372-4 Sequence 4, Appli
29	94	4.4	34303	3	US-09-244-752-4 Sequence 4, Appli
30	94	4.4	34303	3	US-09-245-497-4 Sequence 4, Appli
31	94	4.4	34303	4	US-09-562-919-4 Sequence 4, Appli
32	94	4.4	34382	2	US-08-374-483-6 Sequence 6, Appli
33	94	4.4	35408	3	US-08-973-334-3 Sequence 3, Appli
34	94	4.4	35408	3	US-09-563-869A-3 Sequence 3, Appli
35	94	4.4	35408	3	US-08-549-489-3 Sequence 3, Appli
36	94	4.4	35935	2	US-08-735-609-1 Sequence 1, Appli
37	94	4.4	35935	2	US-08-735-609-1 Sequence 1, Appli
38	94	4.4	35935	3	US-08-379-452-43 Sequence 43, Appl
39	94	4.4	35935	3	US-09-315-372-1 Sequence 1, Appli
40	94	4.4	35935	3	US-09-244-752-1 Sequence 1, Appli
41	94	4.4	35935	3	US-09-245-497-1 Sequence 1, Appli
42	94	4.4	35935	3	US-09-409-670-43 Sequence 43, Appl
43	94	4.4	35935	4	US-09-562-919-1 Sequence 1, Appli
44	93	4.4	651	4	US-09-252-991A-11750 Sequence 11750, A
45	93	4.4	765	4	US-09-252-991A-11829 Sequence 11829, A

#### ALIGNMENTS

##### RESULT 1

US-09-328-111-215/c  
; Sequence 215, Application US/09328111  
; Patent No. 6262333  
; GENERAL INFORMATION:  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astie, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Derti, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/328,111  
; CURRENT FILING DATE: 1999-06-08  
; EARLIER APPLICATION NUMBER: US 60/088,801  
; EARLIER FILING DATE: 1998-06-10  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 215  
; LENGTH: 618  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(618)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-328-111-215



## Alignment Scores:

Pred. No.: 2.17e-13 Length: 618  
Score: 195.00 Matches: 45  
Percent Similarity: 80.00% Conservative: 3  
Best Local Similarity: 75.00% Mismatches: 9  
Query Match: 9.15% Indels: 3  
DB: 3 Gaps: 0

US-09-729-674-2 (1-401) x US-09-328-111-215 (1-618)

QY 345 TrpGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGlnPheAsnVal 364  
DB 515 TGGGGCCCTTNGAATGAGCANTTCCAGCCAGNCTTAAGACGTTCAATGTTCTTAAATGTG.456  
QY 365 ThrThr-LeuGlnLysGlu-LeuTyrrAspPhe-AlaLysGluAsnIleMetAspAsp 383  
DB 455 NCCACCACTCCAGANGAAGCTTGATGACITTTGGNTAAGGAAATATAATGGATGAT 396  
QY 384 GluGlyGluValValGluTyrrValAspLeuLeuGluLeuGluThrSer 401  
DB 395 GAGGAGAGAGTGTGGAATATGTGGATGACCTCTTGAAGTGGAGGAGACCACT 342

## RESULT 2

US-08-961-527-108  
Sequence 108 Application US/08961527  
Patent No. 6420135

## GENERAL INFORMATION:

APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:

## CLASSIFICATION: 424

## PRIOR APPLICATION DATA:

## FILING DATE:

## ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:

## TELEPHONE: (301) 309-8504

## TELEFAX: (301) 309-8512

## INFORMATION FOR SEQ ID NO: 108:

## SEQUENCE CHARACTERISTICS:

## LENGTH: 11309 base pairs

## TYPE: nucleic acid

## STRANDEDNESS: double

## TOPOLOGY: linear

US-08-961-527-108

## Alignment Scores:

Pred. No.: 5.82e-05 Length: 11309  
Score: 140.00 Matches: 91  
Percent Similarity: 36.67% Conservative: 70  
Best Local Similarity: 20.73% Mismatches: 152  
Query Match: 6.57% Indels: 126  
DB: 4 Gaps: 22

US-09-729-674-2 (1-401) x US-08-961-527-108 (1-11309)

QY 13 AlaLeuLeuCysValAlaCysAla-----LeuArgAlaGlyArgAlaClnTyr 28  
DB 4169 GCATACACTGCCAGCTTGTGGAGCCAGCTCCATTTGCTCAAGGACCAAGTGTACGG 4228  
QY 29 GluArgTyrSerPheArgSerPheProArgAspGluLeuMetProLeuGluSerAlaTyr 48  
DB 4229 ACGCATTTTGGCCATAAATCCTTAAAGACGTGATTTTCTTTGAAAATGAAGTCCA 4288  
QY 49 ArgHisAlaLeuAspLysTyr-SerClyGluHisTrpAlaGluSer\*\*GlyTyrLeuGlu 68  
DB 4289 GAACACCTGGCCCAATPAAGGAATCCCTCTATCAGTGGTTGAAAAGAGACAAAGTTCAA 4348  
QY 69 IleSerLeuArgLeuHisArgLeuLeuArgAspSerGluAlaPheCysHisArgAsn--- 87  
DB 4349 TTAGAGTACCCGCTTTAGAACTTAACAGATTGGGATGTATTTGTAATGGCAATCTA 4408  
QY 88 -----CysSerAlaAlaProGln-----ProGluProAlaAlaGly 99  
DB 4409 GCTCTAGAAGTTCACTGTAGTCCCTTGCTCAGAAAGTCTTAAAGAGCGAAGTGAGGC 4468  
QY 100 LeuAlaSer-----TyrProGluLeuArgLeuPheGlyGlyLeuLeuArgArgAlaHis 117  
DB 4469 TATCGTAGTCAGGTTTACCAAGTACTGTGGTTGCTGGGTCAAAACTGTGCTCAAGGAG 4528  
QY 118 CysLeuLysArgCysLysGlnGlyLeuProAlaPheArgGlnSer----- 132  
DB 4529 CGTTTGACTCGCTACAGCAAGGTTTCTTATTATTCAGTCAAAACATGGGCTTTATGTT 4588  
QY 133 -----GlnProSerArgGluValLeu-----AlaAspPhe 142  
DB 4589 TGGGAATTAGACAGGAAAACAAAGTTTAAAGACTCAAAATACCTGATTACAGGATCTC 4648  
QY 143 GlnArgArgGluProTyrLysPheLeuGlnPheAlaTyrPheLysAlaAsn----- 159  
DB 4649 CGCGTAAACTCCATTATCAATCAAGAAATTTCTTATGCTCAAGTAGTTTATTGAA 4708  
QY 160 -----AsnLeuPro-----LysAlaIleAlaAlaHisThrPheLeuLeuLysHisPro 176  
DB 4709 ATATTGCGTCTTCCCTATPAAGAGACAAAATAATCTCATTTTACAGTTCTCGAGACAAG 4768  
QY 177 AspAspGluMetMetLysArgAsnMetAlaTyrTyrLysSerLeuProGlyAlaGluAsp 196  
DB 4769 GACATCTGCTGCTATATCCGCAACAACCTTTATTATCAAAATCTC-----TTT 4816  
QY 197 TyrIleLysAspLeuGluThrLysSerTyrGluSerLeuPheIleArgAlaValArgAla 216  
DB 4817 TGGATGAAGAA---CAAGCAGAAAGCCTATCAAA----- 4846  
QY 217 TyrAsnGlyGluAsnTrpArgThrSerIleThrAspMetGluLeuAlaLeuProAspPhe 236  
DB 4847 ---AAGGAGAAATAATATCCTGACT----- 4867  
QY 237 PheLysAlaPheTyrGluCysLeuAlaCysGluGlySerArgGluIleLysAspPhe 256  
DB 4868 -----TATGGACTG 4876  
QY 257 LysAspPheTyrLeuSerIleAlaAspHisTyrValGluValLeuGluCysLysIleGln 276  
DB 4877 AAGAATGGTAT----- 4888  
QY 277 CysGluGluAsnLeuThrProValIleGlyGly---TyrProValGluLysPheValAla 295  
DB 4889 -----CCACAAATTCGACCAATAGTGGCAAAATTTTCCAGATTGAACAGCTTGACT 4942  
QY 296 ThrMetTyrHisTyrLeuGlnPheAlaTyrTyrLys-----LeuAsnAspLeuLys 312  
DB 4943 AGCTATTATCAGCACTTT---TATACCTATTACCAAAAAAATCTCCTCAAAATGATTGCAA 4999  
QY 313 AsnAlaAlaProCysAlaVal-----SerTyrLeuLeuPheAsp----- 325  
DB 5000 AAGCTTTATCCAGGCTTTTATCAGCAATATTTCTTTGAAAATATGTTAGATAGAAA 5059

```
QY 326 -----GlnAsnAspLysValMetGlnGlnAsnLeuValTyrGlnTyrHisArgAsp 343
Db 5060 GGATGGAGGAACTCTAATGTTATTACAAAGAAATGAATA-----AATGAAAAGAT 5110
QY 344 ThrTrpGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGlnPhePheAsn 363
Db 5111 ACATGGGATCTATCAACG-----ATCTACCCCACTGACGAGCT----- 5149
QY 364 ValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGlu---AsnIleMetAspAsp 382
Db 5150 -----TGGGAGAGAGCCCTTAAAGATTTAACAGACAAATGGACACAGTAGCCAG 5200
QY 383 AspGluGlyGluValValGluTyrValAspAspLeuLeuGluGluGluThrSer 401
Db 5201 TATGAAGGCCATCTCTTGATAGTGGGATACCTACTAGAAATCACTGAATTTTCT 5257

RESULT 3
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
```

```
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
```

LOCATION: (855539)..(855539)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (871619)..(871619)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1084830)..(1084830)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1096846)..(1096846)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1119881)..(1119881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1130881)..(1130881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1310988)..(1310988)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1313224)..(1313224)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349473)..(1349473)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1349491)..(1349491)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1470091)..(1470091)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1569020)..(1569020)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1602912)..(1602912)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1603734)..(1603734)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1637998)..(1637998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1664855)..(1664855)  
OTHER INFORMATION: n equals a, t, c, or g

Alignment Scores:  
Pred. No.: 1.58e+03  
Score: 106.00  
Percent Similarity: 38.49%  
Best Local Similarity: 25.10%  
Query Match: 4.98%  
DB: 4  
Gaps: 12

US-09-729-674-2 (1-401) x US-08-916-421B-1 (1-1664976)

QY 181 MetLysArgAsnMetAlaTyrTyrLysSerLeuProGlyAlaGluAspTyrLysLeuVal 200  
Db 495629 CTAAGAGATGATGGCAATCCAAATGGATTAAACT-GCTCAAGATATATAAG--- 495574  
QY 201 LeuGluThrLysSerTyrGluSerLeuPheIleArgAlaValArgAlaTyrAsnGly--- 219  
Db 495573 -----AAAGAGITTTATCA-----GCCTACTTCGGTTCG 495544  
QY 220 -----GluAsnTrpArgThrSerIleThrAspMetGluLeuAla 232  
Db 495543 GAATTAACAACCTCCAAAATAGAAATCATGGGACTTCATTTAAAGAGTATCATTCAAA 495484  
QY 233 LeuProAspPhePheLysAlaPheTyrGluCysLeuAlaCysGluGlySerArgGlu 252

Db 495483 ATAGCTAAGATTGAGGAGATATTTGACGAA-----GATAGATTG 495445  
QY 253 IleLysAspPheLysAspPheTyrLeuSerIleAlaAspHisTyrValGluValLeuGlu 272  
Db 495444 ATTAAGACATTAAGAGATGCTAAAGAATTTGGT-----ATTGAA 495403  
QY 273 CysLysIleGlnCysGluGlu---AsnLeuThrProValIleGlyGlyTyrProValGlu 291  
Db 495402 TTAAGGTTAGGTTGAAGAGGCAATCTAAGAAA-----GATGGTTATAAACAACAAA 495349  
QY 292 LysPheValAlaThrMetTyrHisTyrLeuGlnPhe-----AlaTyrTyrLys 307  
Db 495348 GTTATGTGCTCAATATACACCAATAAAGATTTTGGAGAATTTGATATACCTAT 495289  
QY 308 LeuAsnAspLeuLysAsnAlaAlaProCysAlaValSerTyrLeuLeuPhe----- 324  
Db 495288 GCTAACAAAAGAAACATTAGCAAGATATGCTATGATATACCTCTTACAAAAGAGAAA 495229  
QY 325 -----AspGlnAsnAspLysValMetGlnAsnLeuValTyrTyrGlnTyrHis 341  
Db 495228 TATCTAAAGACAGAAACATTAAATAAATAACCAAAATTTTATACATTTGAT 495169  
QY 342 ArgAspThrTyrGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGlnPhe 361  
Db 495168 AAG-----TTT 495163  
QY 362 PheAsnValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGluAsnIleMetAsp 381  
Db 495162 ATAAAGAGAAATGCCTTAAGAA-----GGATTGTAAAGAGAAAAATTTGAAGT 495112  
QY 382 AspAspGluGlyGluValValGluTyrValAspAspLeuLeuGluGluThr 400  
Db 495111 ATTGAAGAACAAAA---GTAGATTGTCTATGATATATACTACATCTCAGAAACT 495058

## RESULT 4

US-09-036-987A-1  
Sequence 1, Application US/09036987A  
Patent No. 6143526  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H.  
APPLICANT: Broughton, Mary C.  
APPLICANT: Crawford, Kathryn P.  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Merlo, Donald J.  
APPLICANT: Treadway, Patti J.  
APPLICANT: Turner, Jan R.  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences LLC Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036.987A  
FILING DATE: 09-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stuart, Donald R  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317)337-4816  
TELEFAX: (317)337-4847

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Alignment Scores:
Pred. No.: 31.3 Length: 80161
Score: 102.50 Matches: 72
Percent Similarity: 31.64% Conservative: 34
Best Local Similarity: 21.49% Mismatches: 119
Query Match: 4.81% Indels: 110
DB: Gaps: 15

US-09-729-674-2 (1-401) x US-09-036-987A-1 (1-80161)
QY 73 LeuHisArgLeuLeuArgAsp-SerGluAlaPheCysHisArgAsnGlySerAlaAlaPr 92
Db 19414 ATCGAGCGAGCTTGGCGACATTGGACGGCTGCTCCACAGATTGCCGGCTGCTC
QY 92 oGlnProGlu---ProAlaAlaGlyLeuAla-----101
Db 19474 GGTGTCCGCTGACCGAACCGCGATTGGAAAGCGAGCTGCGTGGTTCCACCGATC 19533
QY 102 -----SerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArgAlaHisCy 118
Db 19534 GCGGTCCTGTCTCCCGCGATC-----CCGGCCGTGTGTGAGCTCGACCATG 19584
QY 118 sLeuLysArgCysLysGlnGlyLeuProAlaPheArgGlnSerGlnPro-----134
Db 19585 CGCGGTCCGCTGACGCTCTGTGTGGCGGATCAGGACCTCGTCCCGGATGCTTCAG 19644
QY 135 -----Se 135
Db 19645 CGCCGACGACGCGACCATTCGCTGATGAGTCGTGGTAAAGCGACGCTGCTCCACCGATC 19704
QY 135 rArgGluValLeuAlaAspPheGlnArgArgGluProTyrLysPheLeuGlnPheAlaTy 155
Db 19705 TAGCGCATACTT-----CGGACCCAGTCTCTTTCCCGGATTAGCGCA 19749
QY 155 r-----Ph 156
Db 19750 GCAGCCCTACTCCCATTTGGCCAGGATTTGGAATAAGCGCTGCTATGTCATCGCCGTT 19809
QY 156 eLysAlaAsnLeuProLysAlaAlaAlaHisThrPheLeuLeuLysHisPr 176
Db 19810 GAGTCCAAACGACTTCCGGCGGCAACATAGTGTGTCAC-----19849
QY 176 oAspAspGluMetMetLysArgAsnMetAlaTyrTyrLysSerLeuProGlyAlaGluAs 196
Db 19850 -----GGCAGGAATGTCACGCGACCATCGAAGATCTTTGGTTCGCGCA 19893
QY 196 pTyrIleLysAspLeuGluThrLysSerTyrGluSerLeuPheIleArgAlaValArgAl 216
Db 19894 CTGTGTTTCCGCGAACGAGTGAATGCGGAGTCC-----GC 19932
QY 216 aTyrAsnGlyGluAsnTrpArgThrSerIleThrAspMetGluLeuAlaLeuProAspPh 236
Db 19933 TCGATCGGG-----TGGCGCGGACCTGACGGTGATCACCCTGTGTTCTCGCGGATT 19986
QY 236 ePhe-----LysAlaPheTyrGluCysLeu-----Al 245
Db 19987 CATGGGAAGATTGCGTGTGGTGTGTTTCCCTCCTGCGCGGATGATTATAGTCGGTACCGC 20046
QY 245 aAlaCysGluGlySerArgGluIleLysAsp-----PheLysAspPheTyrLeuSeril 263
Db 20047 CGCATCGCGGATACCGCGAATTAACTGACGGCTAGTTTCCGCTCTTTCTCTCTGTGT 20106
QY 263 eAlaAspHisTyrValGluValLeuGluCysLysIleGlnCysGluAsnLeuThrPr 283
;

```











Db 484 GCGCTTGTCAACGAGGTAGCGCTCGACGAGCGCTCGCAACTATCGGGGCTGCTCT 543

Qy 211 eArgAlaValArgAlaTyraAsnGlyGluAsnTrpArg 223

Db 544 G---GCCGAGCGCGCGCGCGCGGCGATACCTTGGCGC 577

## RESULT 10

US-09-252-991A-7066/c

; Sequence 7066, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 7066

; LENGTH: 2751

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-7066

## Alignment Scores:

Pred. No.:	0.241	Length:	2751
Score:	100.50	Matches:	46
Percent Similarity:	41.62%	Conservative:	26
Best Local Similarity:	26.59%	Mismatches:	63
Query Match:	4.72%	Indels:	38
DB:	4	Gaps:	9

US-09-729-674-2 (1-401) x US-09-252-991A-7066 (1-2751)

Qy 78 ArgAspSerGluAlaPheCysHisArgAsnCysSerAlaAlaProGlnProGluProAla 97

Db 2668 AATCGTCAGGCGCATTTCCAGCGTCGGGATAGGAGCAATCGGCTTCGCTCA 2609

Qy 98 AlaGly-----LeuAlaSerTyProGluLeuArgPheGlyGlyLeuArg 114

Db 2608 TCGAAGCACAAGGATGTTCCGATGCGCGGACTCTTCGCG-----CCCGTATTCTGCGC 2555

Qy 115 ArgAla-HisCysLeuLeuArgCysLysGlnGlyLeuProAlaPheArgGlnSerGlnPr 134

Db 2554 CGCGCTCTCTGCTG-----TACCAGCGCTGTCAGGCTGCGCGC 2513

Qy 134 oSerArgGluValLeuAlaAspPheGlnArg-----ArgGluProTy 148

Db 2512 CGAAGCGAAATCTCGCGCTGTTCAGCGCTGGCAGCGCGCCCGCGCGCGCGG 2453

Qy 148 rLysPheLeuGlnPheAlaTyPhe-----LysAlaAsnLe 161

Db 2452 CAAGCGCTCGAAGGGTCTATCGCGCGCTGACTATGCGCGCGCGCGCGCGCGG 2393

Qy 161 uProLysAlaIleAlaAlaHisThrPheLeuLeuLysHisProAspAspGluMetMe 181

Db 2392 CGAGCGACCTCTCGCGCAAGCAGCGCTTCTCGCGCGCGCATCCGACTTCGAG----- 2338

Qy 181 tLysArgAsnMetAlaTyTyTyLysSerLeuProGlyAlaGluAsp----- 196

Db 2337 -CAGCGCATGTCGCTCGCGCGAGATCGTCGCGCGCGCGCGCGCGCGCGCGCGG 2279

Qy 197 ----TyriLeLys-----AspLeuGluThrLysSerTyTyLysSerLeuPhe 211

Db 2278 GCCTTCGTCAACGAGGTAGCGCTCGACGAGCGCTGCGCAACTATCCGGGCTGCTCT 2219

Qy 211 eArgAlaValArgAlaTyraAsnGlyGluAsnTrpArg 223

Db 2218 G---GCCGAGCGCGCGCGCGCGGCGGCGATACCTTGGCGC 2185

## RESULT 11

US-09-221-017B-545

; Sequence 545, Application US/09221017B

; Patent No. 6444799

; GENERAL INFORMATION:

; APPLICANT: Ross, Bruce C.

; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

; NUMBER OF SEQUENCES: 1120

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 PAGE MILL ROAD

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/221,017B

; FILING DATE: 23-DEC-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP1182

; FILING DATE: 31-DEC-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP1546

; FILING DATE: 30-JAN-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP2911

; FILING DATE: 09-APR-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/AU98/01023

; FILING DATE: 10-DEC-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Monroy, Gladys H

; REGISTRATION NUMBER: 32,430

; REFERENCE/DOCKET NUMBER: 27340-20021.00

; TELEPHONE: 650-813-5600

; TELEFAX: 650-494-0792

; TELEX: 706141

; INFORMATION FOR SEQ ID NO: 545:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5633 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: UNKNOWN

; ORIGINAL SOURCE:

; ORGANISM: PORPHYROMONAS GINGIVALIS

; NAME/KEY: misc.feature

; LOCATION: 1...5633

US-09-221-017B-545

## Alignment Scores:

Pred. No.:	0.761	Length:	5633
Score:	100.50	Matches:	92
Percent Similarity:	36.20%	Conservative:	68
Best Local Similarity:	20.81%	Mismatches:	165
Query Match:	4.72%	Indels:	118
DB:	4	Gaps:	23

US-09-729-674-2 (1-401) x US-09-221-017B-545 (1-5633)

Qy 31 TyrSerPheArgSerPheProArgAspGluLeuMetProLeuGluSerAla----- 47

Db 31 TyrSerPheArgSerPheProArgAspGluLeuMetProLeuGluSerAla----- 47

Db 1796 TATCTTCTCGGGCGGTGGCAGCGAGAAATCTGGCAAAATCGATTGAGCATTCAGGAC 1855  
Qy TyrArgHisAlaLeuAspLysTyrSerGlyGluHisTrpAlaGluSer\*\*GlyTyrLeu 67  
Db 1856 TACCGTCCGGATGGAGCTGATCGGAACGATGAG-----GGCATGCTG 1900  
Qy GluIleSerLeuArgLeuHisArgLeuLeuArgAspSerGluAlaPheCysHis----- 85  
Db 1901 -----GTGAATCTCACCAGGAGTGTCTACCCAGCAGCAACG-CTATGCCGAAGCAG 1950  
Qy 86 -----ArgAsnCysSerAlaAlaProGlnProGluProAla----- 97  
Db 1951 GGAAGGGGTAGCGGACTGCTCAATCTCTACCCCAAGCAAGCAGCAGATAGTCTCT 2010  
Qy 98 -----AlaGlyLeuAlaSerTyrProGluLeuArgLeuPheGlyGlyLeu 113  
Db 2011 CAGCGCATAGAGTGGCGAGCAGGATAC----- 2040  
Qy 114 ArgArgAlaHisCysLeuLysArgCysLysGlnGly-Leu----- 126  
Db 2041 CGTCGACGCCAT-----TCGGGAAGTGAACGAGGACTTCGGATGAGACTCTCTCTTGC 2094  
Qy 127 -ProAlaPheArgGln-----SerGlnProSerArgGluValLe 139  
Db 2095 TCCTGCTTATCGCAGATGGCATGCTGCACCTCAAGAGCAAGCGCAACGAGGAGCTAT 2154  
Qy 139 uAlaAspPheGlnArgArg-----GluProTyrLysPheLeuGlnPhe----- 153  
Db 2155 GGCAGCCCTGCAGCAAGCAAGTGAATTCGGAGCCGAGCAACTGTCCCACTATCAACCG 2214  
Qy 154 -----AlaTyrPheLysAlaAsnLeuProLysAlaIleAlaAlaHisThrPh 171  
Db 2215 CGGTGTCATCGCTACCACTCGAAGCACTTCGGTGGAGCCATGGACGACTACACCCAGT 2274  
Qy 171 eLeuLeuLysHisProAspAspGluMetMetLysArgAsnMetAlaTyrTyrLysSerLe 191  
Db 2275 CGTTCGCAAAAGCGCAAGCACTGGCAGCTTTCACCGTGTCTTCCTCCGCTCTTA 2334  
Qy 191 uProGly-----AlaGluAspTyr-----IleLysAspLeuGluThrLy 204  
Db 2335 CTGCGTACGTGAACATGCGATCGAGACTTCGATGATGATCATCCCTCTGGAGCCGGA 2394  
Qy 204 SerTyrGluSerLeuPheIleArgAlaValArgAlaTyrAsnGlyGluAsnTrpArgTh 224  
Db 2395 GAACATATCATGCACTCAACCGTGCCTCTCTCCTCAGCAGATCAACGAGAACCGAAA 2454  
Qy 224 rSerIleThrAspMetGluLeuLeuLeu-----ProAspPhePheLysAlaPheTy 241  
Db 2455 AGCCATTACCGACTTCGACAAAGTCTGGACACTATCTCTGACTTTGTCTGGCTACTA 2514  
Qy 241 rGluCysLeuAlaAlaCysGluGlySerArgGluIleLysAspPhe-----Ly 257  
Db 2515 T-----GTCGCTCCCAAGCCAAAGAGCATTTGGCGNACATCTCGGAGCGGACG 2565  
Qy 257 sAspPheTyrLeuSerIle----- 263  
Db 2566 GGACTACTCGCGCGCTTTGACATAGAAAAAGCGGCACAGACAGCAACAAAAACGAA 2625  
Qy 264 ----AlaAspHisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsnLeuTh 282  
Db 2626 AAAAGACAGCTCTTCGACTTCACCAAGAGACAGCTGAGGATACCGACGAAACG----- 2680  
Qy 282 rProValIleGlyGlyTyrProValGluLysPheValAlaThrMetTyrHisTyrLeuGl 302  
Db 2681 -----ATAGAGAAGTTC-----AACCTGCT 2700  
Qy 302 nPheAlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAlaProCysAlaValSerTyrLe 322  
Db 2701 CDTGATATCCGAGAAATCTCCGAGAGCGGCACAGCTACTCCAGTCGGATCGTGGCGC 2760  
Qy 322 uLeuPheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyrTyrGlnTyrHisAr 342  
Db 2761 GATTGAGGACACGATGTAGAGGTATACAGCCCTCGCGGCTATTTCGTCCTCAGCTACTAGA 2820

Qy 342 gAspThrTrpGlyLeuSerAspGluHisPheGlnProArg-----ProGluAlaVa 359  
Db 2821 GCAG-----TCTCTTCGAAAGAGTGGCGCGCTACTACTCTCGAAGCCAT 2868  
Qy 359 lGlnPhePheAsn---ValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGluAs 378  
Db 2869 TGCTGCTCTCAACGCAAAAGGTTCTGCCAAGGACTCAAAACCGATCAACCGTGAGT 2928  
Qy 378 nileMetAspAspAspGluGlyGluValValGluTyrValAspAspLeuLeuGluLeuGl 398  
Db 2929 AGCTCTGAGTCAAGAG-----CAGGTGGCATATCAGCAACAGGACATAACGAAATATC 2982  
Qy 398 uGlu 399  
Db 2983 CGAG 2986

## RESULT 12

US-09-252-991A-6950  
; Sequence 6950, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107195.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6950

; LENGTH: 2316

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6950

## Alignment Scores:

Pred. No.: 0.24 Length: 2316  
Score: 99.50 Matches: 92  
Percent Similarity: 35.49% Conservative: 56  
Best Local Similarity: 22.06% Mismatches: 129  
Query Match: 4.67% Indels: 140  
Gaps: 25

US-09-729-674-2 (1-401) x US-09-252-991A-6950 (1-2316)

Qy 1 MetGluProGlyArgArg-----GlyAlaAlaAlaLeuAlaLeuLeuCys 16

Db 510 CTGACCAAGCGCGTCGACGATCGGCAAGGACGCGCATGCT----- 551

Qy 17 ValAlaCysAla-----LeuArgAlaGlyArgAlaGlnTyrGluArgTyrSerPhe 33

Db 552 GTCTCAGCGCGTGATCCCGGATTCAGGCGGTGCGGCAATATCCT-----GAAAT 605

Qy 34 ArgSerPheProArgAspGluLeuMetProLeuGluSerAlaTyrArg-----HisAla 51

Db 606 CGAAGTCTGCACCGCGCGAT-----GAACAAGCAGTTCGAAGCCTCCGCTCG 653

Qy 52 LeuAspLysTyrSerGlyGluHis---TrpAlaGluSer---GlyTyrLeuGluLeuSer 70

Db 654 GTCCGCGCGATACAGCGAGTCGATCCACGCGAAAGTCGTCACACCGCGTGAATGA 713

Qy 71 LeuArgLeuHisArgLeuLeuArgAspSerGluAlaPheCysHisArgAsnCysSerAla 90

Db 714 CGTCGATATCACCGA-----CCGATTCCGCTCTATGGCGCATCAACAT----- 758

Qy 91 AlaProGlnProGluProAlaAlaGlyLeuAlaSerTyrProGluLeuArgLeuPheGly 110

Db 759 -----GGTCCAGAAATACAAAGCGGAGGTACACCGTACACCGTACCAATGGAAGC 812

QY 111 -----GlyLeuLeuAargAlaHisCysLeuLysArgCysLysGlnGlyLeuProAla 128  
Db 813 CAACGCAATTCCTACATGAACCCGATGTGTGGCGC----- 851  
QY 129 PheArgGlnSerClnProSerArgGlu-----ValLeuAlaAspPheGlnArgArgGluPro 147  
Db 852 -----CCTCACATGGAAAGCAATCTCGGAGGCTTCAGAAAGGAAACAGCAA 896  
QY 148 Tyr-LysPheLeuGlnPheAlaTyrPheLysAlaAsnLeuProLysAlaIleAlaAl 167  
Db 897 CACCAAGGTCATCAAA-----GACCCATCTCGC 926  
QY 167 alaHisThrPheLeuLeuLysHisProAspAspGluMet-----MetLysArgAs 184  
Db 927 GGGCCACGCGAAAGCAATTCAGAGATGATCCAGGACATCCGCGCGACAGCCACAGAT 986  
QY 184 nMetAlaTyrTyrLysSerLeuProGlyAlaGluAspTyrIleLysAspLeuGluThrLy 204  
Db 987 CGTCGCGGGCAAAAGCCGGCTCCGGC-----GTGATCGATGCCGAGCCAA 1034  
QY 204 sSerTyrGluSerLeuPheIleArgAlaValArgAlaTyrAsnGlyGluAsnTrpArgTh 224  
Db 1035 CAAATACGAGAACTCGCTCGCGCCGCTAAAGCTGCGCTCACAACG----- 1086  
QY 224 rSerIleThrAspMetGluLeuAlaLeuProAspPhePheLysAlaPheTyrGluCysLe 244  
Db 1087 -----AGTGATCGGGCTCGAGTCACTTCCTCAAGGCC----- 1119  
QY 244 uAlaAlaCysGluGlySerArgGluIleLysAspPheLysAspPheTyrLeuSerIleAl 264  
Db 1120 -----GCCAAGGAT----- 1128  
QY 264 aAspHisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsnLeuThrProVa 284  
Db 1128 ----- 1128  
QY 284 lIleGlyGlyTyrProValGluLysPheValAlaThrMetTyrHisTyrLeuGlnPheAl 304  
Db 1129 -----GTGTTCGCGGAACCTGGTACACCATC----- 1167  
QY 304 aTyrTyrLysLeuAspLeuLys-----AsnAlaAlaProCysAlaValSe 320  
Db 1168 -----GTCAAGATGAACGAGCTGTAGTCAGTCCAGCTGAACGGCTTCCGCGCGAGAAC 1223  
QY 320 rTyrLeuLeuPheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyrGlnTy 340  
Db 1224 CATTCGATCTCGCAAGAGAGACCAAGGAGACACTCCAGAC----- 1266  
QY 340 rHisArgAspThrTrpGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGl 360  
Db 1267 -TATGCTGACGTGATGCTTACCGCTGACGAATATGCAAGCACCGCGTGGACTCAGTT-- 1323  
QY 360 nPhePheAsnValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGluAsnIleMe 380  
Db 1324 -----AGGAACACTAC---TATCGGAAACCAACATCG 1355  
QY 380 tAspAspAspGluGlyGlu---ValValGluTyrValAspAspLeuLeu 395  
Db 1356 AGAGCCCGGTGACGCGCAAGGGCGCTGGGAGTACGTCAAGAGATGATG 1404  
RESULT 13  
US-09-729-674-2 (1-401) x US-09-729-674-2 (1-963)  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6979  
; LENGTH: 963  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-729-674-2 (1-401) x US-09-729-674-2 (1-963)  
Alignment Scores:  
Pred. No.: 0.115 Length: 963  
Score: 97.00 Matches: 39  
Best Local Similarity: 44.53% Conservative: 22  
Query Match: 28.47% Mismatches: 43  
Indels: 33  
Caps: 7  
US-09-729-674-2 (1-401) x US-09-729-674-2 (1-963)  
QY 111 GlyLeuLeuAargAlaHisCysLeuLysArgCysLysGlnGlyLeuProAlaPheAr 130  
Db 66 GGTATTCGCGCGCGCTCTTCGCTG-----TACCCAGCGGTGGT 107  
QY 130 gGlnSerGlnProSerArgGluValLeuAlaAspPheGlnArg----- 144  
Db 108 CCAGGCTCGCGCGCAAGCGAATCTTCGCGCTGTTCCAGCGTGGCAGCGCCGCCACGG 167  
QY 145 -ArgGluProTyrLysPheLeuGlnPheAlaTyrPhe-----Ly 157  
Db 168 GCAGCGGAGCGCGCAAGCGCTGGAAGGGTCTATGCGCGCGCTGACTACTATGCGCG 227  
QY 157 sAlaAsnAsnLeuProLysAlaIleAlaAlaHisThrPheLeuLeuLysHisProAs 177  
Db 228 CGCCAGACCGCGCGAGCGCACCTCTCCGCCAAGCAGGCGCTTTCGCGCGCATCCCGA 287  
QY 177 pAspGluMetMethLysArgAsnMetAlaTyrTyrLysSerLeuProGlyAlaGluAsp-- 196  
Db 288 CTTCGAG-----CAGCGCATCGTGTGCGCGCGAGATCGTCCGCGGAGCAGGAAAG 341  
QY 197 -----TyrIleLys-----AspLeuGluThrLysSerTyrGl 207  
Db 342 CCGTACGAGTGGCGCTTCGTCAAGCAGGATACGCTCGAGGAGCGCTGGCGCACTATCC 401  
QY 207 uSerLeuPheIleArgAlaValArgAlaTyrAsnGlyGluAsnTrpArg 223  
Db 402 GGGCTCTCTCTG---CCCGAGCGCGCGCGCGCGCGCATACCTGGCGC 447  
RESULT 14  
US-09-729-674-2 (1-401) x US-09-729-674-2 (1-963)  
; Sequence 15346, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15346  
; LENGTH: 2283  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-729-674-2 (1-401) x US-09-729-674-2 (1-963)  
Alignment Scores:  
Pred. No.: 0.458 Length: 2283  
Score: 97.00 Matches: 87  
Percent Similarity: 30.24% Conservative: 37

QY 111 -----GlyLeuLeuAargAlaHisCysLeuLysArgCysLysGlnGlyLeuProAla 128  
Db 813 CAACGCAATTCCTACATGAACCCGATGTGTGGCGC----- 851  
QY 129 PheArgGlnSerClnProSerArgGlu-----ValLeuAlaAspPheGlnArgArgGluPro 147  
Db 852 -----CCTCACATGGAAAGCAATCTCGGAGGCTTCAGAAAGGAAACAGCAA 896  
QY 148 Tyr-LysPheLeuGlnPheAlaTyrPheLysAlaAsnLeuProLysAlaIleAlaAl 167  
Db 897 CACCAAGGTCATCAAA-----GACCCATCTCGC 926  
QY 167 alaHisThrPheLeuLeuLysHisProAspAspGluMet-----MetLysArgAs 184  
Db 927 GGGCCACGCGAAAGCAATTCAGAGATGATCCAGGACATCCGCGCGACAGCCACAGAT 986  
QY 184 nMetAlaTyrTyrLysSerLeuProGlyAlaGluAspTyrIleLysAspLeuGluThrLy 204  
Db 987 CGTCGCGGGCAAAAGCCGGCTCCGGC-----GTGATCGATGCCGAGCCAA 1034  
QY 204 sSerTyrGluSerLeuPheIleArgAlaValArgAlaTyrAsnGlyGluAsnTrpArgTh 224  
Db 1035 CAAATACGAGAACTCGCTCGCGCCGCTAAAGCTGCGCTCACAACG----- 1086  
QY 224 rSerIleThrAspMetGluLeuAlaLeuProAspPhePheLysAlaPheTyrGluCysLe 244  
Db 1087 -----AGTGATCGGGCTCGAGTCACTTCCTCAAGGCC----- 1119  
QY 244 uAlaAlaCysGluGlySerArgGluIleLysAspPheLysAspPheTyrLeuSerIleAl 264  
Db 1120 -----GCCAAGGAT----- 1128  
QY 264 aAspHisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsnLeuThrProVa 284  
Db 1128 ----- 1128  
QY 284 lIleGlyGlyTyrProValGluLysPheValAlaThrMetTyrHisTyrLeuGlnPheAl 304  
Db 1129 -----GTGTTCGCGGAACCTGGTACACCATC----- 1167  
QY 304 aTyrTyrLysLeuAspLeuLys-----AsnAlaAlaProCysAlaValSe 320  
Db 1168 -----GTCAAGATGAACGAGCTGTAGTCAGTCCAGCTGAACGGCTTCCGCGCGAGAAC 1223  
QY 320 rTyrLeuLeuPheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyrGlnTy 340  
Db 1224 CATTCGATCTCGCAAGAGAGACCAAGGAGACACTCCAGAC----- 1266  
QY 340 rHisArgAspThrTrpGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGl 360  
Db 1267 -TATGCTGACGTGATGCTTACCGCTGACGAATATGCAAGCACCGCGTGGACTCAGTT-- 1323  
QY 360 nPhePheAsnValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGluAsnIleMe 380  
Db 1324 -----AGGAACACTAC---TATCGGAAACCAACATCG 1355  
QY 380 tAspAspAspGluGlyGlu---ValValGluTyrValAspAspLeuLeu 395  
Db 1356 AGAGCCCGGTGACGCGCAAGGGCGCTGGGAGTACGTCAAGAGATGATG 1404  
RESULT 13  
US-09-729-674-2 (1-401) x US-09-729-674-2 (1-963)  
; Sequence 6979, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18

Best local Similarity: 21.22% Mismatches: 125  
Query Match: 4.55% Indels: 161  
DB: 4 Gaps: 17

US-09-729-674-2 (1-401) x US-09-252-991A-15346 (1-2283)

QY 22 ArgAlaGlyArgAlaGlnTyrGlu-----ArgTyrSerPheArgSerPhe 36  
DB 1087 CGGCTGGAAGACAGAGCGCTTCCCGTACTGTCAAGACGCGCTCGTGGCGCGAATT 1146  
QY 37 ProArgAspGluLeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSer 56  
DB 1147 CCCGCTGATGCGTGCAGCCCTGATGATCCGCGCACCG-----1185  
QY 57 GlyGluHisTrpAlaGluSer\*\*GlyTyrLeuGluHisLeuLeuArgLeuHisArgLeu 76  
DB 1186 -----CGCGCTGTTCCGCTTC 1200  
QY 77 LeuArgAspSerGluAlaPheCysHisArgAsnCysSerAlaAlaProGlnProGluPro 96  
DB 1201 CTTCCGCGCTCA-----CCCGAGCTT 1221  
QY 97 AlaAlaGlyLeuAlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArgAla 116  
DB 1222 CGAGTGGCCCTGGAGCGCAGCTGACCGAAGTCTCCAGGGCGCGACCTTCGAGGCT 1281  
QY 117 His-----CysLeuLysArgCysLysGlnGlyLeu 126  
DB 1282 GAACGACCTGCCCGCCAGCGACCTTCGAAAGCATCTCGCTGACCGAGCGGAACAACCTTCGT 1341  
QY 127 ProAlaPheArgGlnSerGlnProSerArgGluValLeuAlaAspPheGlnArg-ArgG1 146  
DB 1342 CGAGCATTTTCATGATTCCAG---CGGCTGGTGTCTTGGGCTTCTTCAGCGCAGGC 1398  
QY 146 uProTyrLysPheLeuGlnPheAlaTyrPheLysAlaAsnLeuProLysAlaLeu 166  
DB 1399 CGATTTCCGCTGCGAGTGGACTTCACCGCGAGCGCGAGGCTGCCAAGCGCGAGGA 1458  
QY 166 alaAlaHisThrPheLeuLeuLysHisProAspAspGluMetMetLysArgAsn-----184  
DB 1459 AGCCCGACCCCTGTTCCGACCTCTC-----GAACCATGGCGCAAGAGGCTTA 1506  
QY 185 -MetAlaTyrTyrLysSerLeu-----ProGly-----193  
DB 1507 CATGCGGTGTACGAGACCTCGGCGCTACGCGCTCGCGATCTCTGTGCGCGGCTACTC 1566  
QY 194 -----AlaGluAspTyrLysLeuGluThrLysSerTyrGluSerLe 209  
DB 1567 GGAATCTATCCGTTGGAAGACCTGATCTGGGCAACCAACCAAGCGCTCGCG-----1621  
QY 209 uPheIleArgAlaValArgAlaTyrAsnGlyGluAsnTrpArgThrSerIleThrAspMe 229  
DB 1622 -----TTCGCGAGGACATCTCAACCT 1644  
QY 229 tGluLeuAlaLeuProAspPhePheLysAlaPheTyrGluCysLeuAlaLysGluG1 249  
DB 1645 GCACCGCTCGACGACGCGCGCTCGGGCGCTTGTGGAGCGCTGGAAGACTGC-----1699  
QY 249 ySerArgGluLysAspPheLysAspPheTyrLeuSerIleAlaAspHisTyr-----267  
DB 1700 -----GAGGTGGACGACTACCGCATCACACGCTGATCGCGCTGAGTTCGACGA 1752  
QY 268 -----ValGluValLeuGluCysLysIleGlnCysGluGluAsnLe 281  
DB 1753 CAACACGCTGCGGCCCACTGACCATCTCTCGAGCTGAAGTG-----1795  
QY 281 uThrProValIleGlyGlyTyrProValGluLysPheValAlaThrMetTyrHisTyrLe 301  
DB 1796 -----CTCATCGGC-----1804  
QY 301 uGlnPheAlaTyrTyrLysLeuAsnAspLeuLysAlaAlaProCysAlaValSerTyr 321  
DB 1805 -----CTCGCCCTGAAGCGCTTGAAGAGCGCGCAAGAGCGGTGGAGGCTTCTTCCAGTA 1860

QY 321 rLeuLeuPheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyrGlnTyrHi 341  
DB 1861 C-----AAGGACACAGCGCTCGAGCGCGCTGTTCTAC-----1894  
QY 341 sArgAspThrTrpGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGlnPh 361  
DB 1895 -----CAGGC 1899  
QY 361 ePheAsnValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGluAsnIleMetAs 381  
DB 1900 GCTGAACGTGGTGGCTG-----GAGGTGCTCT 1926  
QY 381 pAspAspGluGlyGluValValGluTyr 390  
DB 1927 CGACGACGAGCTGGAGATCGCGACTAC 1954

RESULT 15  
US-09-252-991A-15430/c  
; Sequence 15430, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15430  
; LENGTH: 2463  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15430

Alignment Scores:  
Pred. No.: 0.517 Length: 2463  
Score: 97.00 Matches: 87  
Percent Similarity: 30.24% Conservative: 37  
Best Local Similarity: 21.22% Mismatches: 125  
Query Match: 4.55% Indels: 161  
DB: 4 Gaps: 17

US-09-729-674-2 (1-401) x US-09-252-991A-15430 (1-2463)

QY 22 ArgAlaGlyArgAlaGlnTyrGlu-----ArgTyrSerPheArgSerPhe 36  
DB 1204 CGGCTGGAAGACAGAGCGCTTCCCGTACTGTCTCAAGGACGCTCGTGGCGCGAATT 1145  
QY 37 ProArgAspGluLeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSer 56  
DB 1144 CCGGTGATGCTGCTGACCGCTGATGAATCCGCGCACCG-----1106  
QY 57 GlyGluHisTrpAlaGluSer\*\*GlyTyrLeuGluHisLeuLeuArgLeuHisArgLeu 76  
DB 1105 -----CGGCTGTTCCGCTTC 1091  
QY 77 LeuArgAspSerGluAlaPheCysHisArgAsnCysSerAlaAlaProGlnProGluPro 96  
DB 1090 CTTCCGCGCTCA-----CCCGAGCTT 1070  
QY 97 AlaAlaGlyLeuAlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArgAla 116  
DB 1069 CGAGTGGCCCTGGAGCGCAGCTGACCGAAGTGTCTCAGGGCGCGAGCTTCGAGGCT 1010  
QY 117 His-----CysLeuLysArgCysLysGlnGlyLeu 126  
DB 1009 GAACGACCTGCCCCAGCGACCTTCGAAAGCATCTGCGCTGACCGCGCGCAACAACCTTCGT 950

[illegible]

Search completed: January 30, 2004, 08:13:11  
Job time : 1055 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 30, 2004, 03:51:33 ; Search time 2712 Seconds  
(without alignments)  
3593.693 Million cell updates/sec

Title: US-09-729-674-2  
Perfect score: 2130  
Sequence: 1 MEPRGRGAALLALLCVACA.....DDEGEVVEYVDLLELEETS 401

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPFO\_spool/US09729674/runat\_29012004\_095314\_12897/app\_query.fasta\_1.583  
-DB=EST -OPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -L-SET=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTWT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09729674 @CGN 1 1 2810 runat\_29012004\_095314\_12897 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT\_DSBLCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_vrn.\*  
22: em\_gss\_nam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rtd.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1893	88.9	1656	11	AK047506	Mus muscu
2	1868	87.6	1852	11	AK017797	Mus muscu
3	1791.5	84.1	1201	9	AL545483	AL545483
4	1684	79.1	1201	9	AL556976	AL556976
5	1680.5	78.9	1075	9	AL546910	AL546910
6	1663	78.1	1148	13	BU902192	AGENCOURT
7	1618.5	76.0	1201	13	BU902192	AGENCOURT
8	1610	75.6	1201	9	AL517455	AL517455
9	1600	75.1	1201	9	AL541167	AL541167
10	1584	74.4	1201	9	AL541167	AL541167
11	1522.5	71.5	1201	9	AL543090	AL543090
12	1498	70.3	1201	9	AL554750	AL554750
13	1468	68.9	1430	11	AK007624	Mus muscu
14	1426.5	67.0	846	13	BU938140	AGENCOURT
15	1419	66.6	1109	12	BM550012	AGENCOURT
16	1413.5	66.4	1201	9	AL558186	AL558186
17	1406	66.0	1152	9	AL550966	AL550966
18	1401.5	65.8	944	13	BU956662	AGENCOURT
19	1397	65.6	939	13	BU956662	AGENCOURT
20	1394.5	65.5	1083	12	BM920687	AGENCOURT
21	1369	64.3	774	9	AL598447	AL598447
22	1354	63.6	846	12	BU934401	AGENCOURT
23	1352	63.5	1040	9	AL558966	AL558966
24	1342	63.0	899	13	BU937797	AGENCOURT
25	1341.5	63.0	962	14	CD516948	AGENCOURT
26	1337.5	62.8	904	13	BU931550	AGENCOURT
27	1325.5	62.2	1103	13	BQ333085	AGENCOURT
28	1325	62.2	1200	13	BU902214	AGENCOURT
29	1298	60.9	878	13	BU183287	AGENCOURT
30	1285.5	60.4	945	13	BQ677531	AGENCOURT
31	1285	60.3	827	12	BU760650	AGENCOURT
32	1277	60.0	875	14	CD359485	AGENCOURT
33	1277	60.0	912	13	BU175306	AGENCOURT
34	1262	59.2	1201	13	BU937872	AGENCOURT
35	1261	59.2	887	13	BU718961	AGENCOURT
36	1246	58.5	837	13	BQ428356	AGENCOURT
37	1244.5	58.4	782	10	BE877856	AGENCOURT
38	1241	58.3	713	10	BE746070	AGENCOURT
39	1233	57.9	846	13	BU189067	AGENCOURT
40	1228	57.7	707	10	BG747825	AGENCOURT
41	1215.5	57.1	1201	13	BU946004	AGENCOURT
42	1214	57.0	949	10	BG323659	AGENCOURT
43	1211	56.9	754	14	CA322247	AGENCOURT
44	1209	56.8	943	13	BU539494	AGENCOURT
45	1198.5	56.3	1088	10	BF025893	AGENCOURT

ALIGNMENTS

RESULT 1  
AK047506  
LOCUS  
DEFINITION  
AX047506  
Mus musculus 1656 bp mRNA linear HTC 05-DEC-2002  
Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length  
enriched library, clone:B930079M13 product:cartilage associated  
protein, full insert sequence.  
ACCESSION  
AX047506  
VERSION  
AX047506.1 GI:26338809  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Muridae; Murinae; Mus.









```

DEFINITION AL545483 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
ACCESSION clone CS0DI015YK22 5-PRIME, mRNA sequence.
VERSION AL545483
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT
    On Feb 15, 2001 this sequence version replaced gi:12877964.
    Contact: Genoscope
    Genoscope - Centre National de Sequencage
    BP 191 91006 EVRY cedex - France
    Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
    Library was constructed by Life Technologies, a division of
    Invitrogen. This sequence belongs to sequence cluster 2233.f For
    more information about this cluster, see
    http://www.genoscope.cns.fr/
    cgi-bin/cluster.cgi?seq=CS0DI015BF11QI&cluster=2233.f. Contact :
    Feng Liang Email : fliang@lifetech.com URL :
    http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
    Faraday Avenue Genoscope sequence ID : CS0DI015BF11QP1.
    Location/Qualifiers
        1..1201
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DI015YK22"
            /issue_type="PLACENTA COT 25-NORMALIZED"
            /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
            /note="1st strand cDNA was primed with a NotI-oligo(dt)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and cloned into the Not I and EcoR V
            sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      256 a   363 c   321 g   242 t   19 others
ORIGIN
Alignment Scores:
Pred. No.:          1,6e-192           Length:       1201
Score:             1791.50              Matches:       347
Percent Similarity: 95.07%               Conservative:   0
Best Local Similarity: 95.07%             Mismatches:     18
Query Match:       84.11%                 Indels:         3
DB:                9                      Gaps:           0

US-09-729-674-2 (1-401) x AL545483 (1-1201)

Qy      1  MetGluProGlyArgArgGlyAlaAlaLeuAlaLeuAlaLeuCysValAlaCysAla 20
Db      106 ATGCAGCCGGCGCGCGCGGGCGCGCGCTGTAGCGTGCTGTCGTGGCTGCGCG 165
Qy      21  LeuArgAlaGlyArgAlaClnTyrcluhargTyrsPheArgSerPheProArgAspGlu 40
Db      166 CTGCC-CGCCGGCGGCC-CAATACGAACGCTACAGCTTCGCGACTTTCCCACGGGACGAG 223
Qy      41  LeuMetProLeuGluSerAlatyrArgHisalaleuAeplytvsrSerglyGluHistyp 60
Db      224 CTGATCCGCTCGAGTCGGCTACC CGGCACCGCGCTGGACAAGTACACGGCGCGAGCTGG 283
Qy      61  AlaGluSer***GlyTyrLeuGluileserleuArgleuHisArgLeuLeuArgAspSer 80
Db      284 GCCGAGAGCGTGGGCTTACCTGGAGATCAGCTCGGGCTGCACCGCTTGTCTGCAGCAGC 343
Qy      81  GluAlapheCyehisargasnCysserlaalaProGlnProGluProAlaAlaGlyLeu 100
Db      344 AG-GCCTTCTGCCACCGCAACTGACGCCGCGCGCGCGCGCGCGCGCGCGCGCTC 402
Qy      101 AlaserTyrProGluLeuArgLeuPheGlyGlyLeuLeuArgArgAlaHisCysLeuLys 120

```

Bp 191 91006 EVRY cedex - France  
Email: [segrefgenoscope.cns.fr](mailto:segrefgenoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 223.f For more information about this cluster, see

Alignment Scores:	
Pred. NO.:	2.66e-180
Score:	1684.00
Percent Similarity:	79.10%
Best Local Similarity:	96.23%
Query Match:	79.06%
DB:	9
Length:	1201
Matches:	332
Conservative:	3
Mismatches:	10
Indels:	5
Gaps:	0



```
Db 123 CGCTCGAGTCGGCTTACCGGACGCGCTGCGCAAGTACAGCGGCGAGCACTGGCGCGAG 182
QY 63 Ser***GlyTyrLeuGluLeuSerLeuArgLeuHisArgLeuLeuArgAspSerGluAla 82
Db 183 AGCGTGGGTACTCTGAGATACGCTCGCGCTGCGCGCTTCTGCTGCGGACAGCGGCC 242
QY 83 PheCysHisArgAsnCysSerAlaAlaProGlnProGluProAlaAlaGlyLeuAlaSer 102
Db 243 TTCTGCCACCGCAACTGACGCGCGCGCGCGCGAGCGCGCGCGCGCGCGCTCGCCAGC 302
QY 103 TyrProGluLeuArgLeuPheGlyLeuLeuArgArgAlaHisCysLeuArgCys 122
Db 303 TATCCGAGCTGGCTCTTGGGGGCGCTGCTGCGCGCGCGCGCGCTGCTCAAGCGCTGC 362
QY 123 LysGlnGlyLeuProAlaPheArgGlnSerGlnProSerArgGluValLeuAlaAspPhe 142
Db 363 AAGCAGGGCTGCGACGCTTCCGCCAGTCCCGAGCCCGCGAGGAGTGTGCGCGACTTC 422
QY 143 GlnArgArgGluProTyrLysPheLeuGlnPheAlaTyrPheLysAlaAsnAsnLeuPro 162
Db 423 CAGCGCGCGAGCGCTTACAGTTCCTGCGAGTTCGCTTACTTCAAGGCAATATCTCCCC 482
QY 163 LysAlaIleAlaAlaAlaHisThrPheLeuLeuLysHisProAspAspGluMetMetLys 182
Db 483 AAGCCATCGCGCTGCTCACACCTTCTTCTGAGCATCTCTGATGACGAATGATGAG 542
QY 183 ArgAsnMetAlaTyrTyrLysSerLeuProGlyAlaGluAspTyrLysAspLeuGlu 202
Db 543 AGGAACATGCGATATATTAAGAGCTGCTGCTGCGGAGGACTACATTAAGACCTGGAA 602
QY 203 ThrLysSerTyrGlnSerLeuPheIleArgAlaValArgAlaTyrAsnGlyGluAsnTrp 222
Db 603 ACCAGTCATATGAAGCTTTCATCCGAGGAGTCCGGGCGATCAACGGTGAGAACTGG 662
QY 223 ArgThrSerIleThrAspMetGluLeuAlaLeuProAspPhePheLysAlaPheTyrGlu 242
Db 663 AGAACATCCATCACAGACATGAGCTGGCGCTTCCCGACTTCTTCAAGCCCTTTACGAG 722
QY 243 CysLeuAlaAlaCysGlnGlySerArgGluIleLysAspPheLysAspPheTyrLeuSer 262
Db 723 TGCTCGCGCTGCGAGGCTTCCAGGAGATCAAGGACTTCAAGGATTTTACCTTTCC 782
QY 263 IleAlaAspHisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsnLeuThr 282
Db 783 ATAGCAGATCATATCTAGAGTTCGGAATGCAATATACATGATGAGAGAACCTCACC 842
QY 283 ProValIleGlyTyrProValGluLysPheValAlaThrMetTyrHisTyrLeuGln 302
Db 843 CCAAGTTAATAGAGGCTATCCGTTGAGAAATTTGGCTTACCATGAATCATTAATTGCGAG 902
QY 303 PheAlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAla-ProCysAlaValSerTyrLe 322
Db 903 TTTCCTATTATAGTTGACGACTGAAATATGCGCGCGCGCTGTTGGAGCGCAGTATCG 962
QY 322 uLeuPheAspGlnAsnAspLysValMetGln-GlnAsnLeuValTyrGlnTyrHis 341
Db 963 GCCTCTGTATCAAGAAAGACAGGGTCTCGCAGCCAACTGCGGGGTATTACCGCGCTCCCC 1022
QY 342 ArgAspThr---TrpGlyLeuSerAsp-GluHisPheGlnProAspProGlu 357
Db 1023 CCGGGGACCTTTGGGGCGCTCGGAGTGACAACTT---CCACGCCCAAAA 1071

RESULT 7
BX425905 1201 bp mRNA linear EST 15-MAY-2003
LOCUS
DEFINITION BX425905 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
CS0DA004YM07 5-PRIME, mRNA sequence.
ACCESSION BX425905
VERSION BX425905.1 GI:30774508
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished
CONTACT: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2233.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0AA004AG04QP1&cluster=2233.f. Contact :
Feng Liang Email : fliang@lifetech.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0AA004AG04QP1.

FEATURES
source
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DA004YM07"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/notes="Vector: pCMVSPORT6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 250 a 334 c 314 g 248 t 55 others
ORIGIN
Alignment Scores:
Pred. No.: 7.4e-173 Length: 1201
Score: 1618.50 Matches: 325
Percent Similarity: 93.93% Conservative: 0
Best Local Similarity: 93.93% Mismatches: 20
Query Match: 75.99% Indels: 5
DB: 13 Gaps: 1
US-09-729-674-2 (1-401) x EX425905 (1-1201)
QY 1 MetGluProGlyArgArgGlyAlaAlaLeuLeuAlaLeuLeuCysValAlaCysAla 20
Db 37 ATGAGCGCGGGCGCGCGGGGGCGCGCGCTGCTAGCGCTGCTGGCTGGCGG 96
QY 21 LeuArgAlaGlyArgAlaGlnTyrGluArgTyrSerPheArgSerPheProArgAspGlu 40
Db 97 CTGCGCGCGGGCGGCC-CAATACGAACGCTACAGCTTCCGAGCTTCCACGGGACGAG 155
QY 41 LeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGlyGluHisTyr 60
Db 156 CTGATGCGCTCGAGTCGGCTTACCGCACGCGCTGGAAGTACAGCGCGGACGACTGG 215
QY 61 AlaGluSer***GlyTyrLeuGluIleSerLeuArgLeuHisArgLeuLeuArgAspSer 80
Db 216 GCGAGAGCGTGGGCTTACCTGGAGATCAGCGCTGCGGCTGCACCGCTGCTGCGCGACA 275
QY 81 GluAlaPheCysHisArgAsnCysSerAlaAlaProGlnProGluProAlaAlaGlyLeu 100
Db 276 G--GCCCTTCTGCCACCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 333
QY 101 AlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArgArgAlaHisCysLeuLys 120
Db 334 GCCAGCTATCCGAGCTGCGCTCTTGGGGGCGCTGCTGCGCGCGCGCACTGCCCTCAAG 393
QY 121 ArgCysLysGlnGlyLeuProAlaPheArgGlnSerGlnProSerArgGluValLeuAla 140
Db 394 CGCTGCAAGCAGGGCGCTGCGAGCGCTTCCGCGAGTTCAGCGCCGCGGAGGCTGGCG 453
QY 141 AspPheGlnArgArgGluProTyrLysPheLeuGlnPheAlaTyrPheLysAlaAsnAsn 160
```















```

RESULT 14
BU838140
LOCUS
DEFINITION AGENCOURT_8071749 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6088309
5', mRNA sequence.
ACCESSION BU838140
VERSION BU838140.1 GI:24022535
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 846)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTP/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2326 row: e column: 14
High quality sequence stop: 664.
FEATURES
source
location/Qualifiers
1..846
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6088309"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC_112"
/note="Organ: skin; Vector: pOTS7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."
BASE COUNT 199 a 255 c 215 g 177 t
ORIGIN
Alignment Scores:
Pred. No.: 2,85e-151 Length: 846
Score: 1426.50 Matches: 269
Percent Similarity: 96.11% Conservative: 3
Best Local Similarity: 95.05% Mismatches: 2
Query Match: 66.97% Indels: 9
DB: 13 Gaps: 1
US-09-729-674-2 (1-401) x BU838140 (1-846)
QY 66 TyrLeuGluIleSerLeuArgLeuHisArgLeuLeuArgAspSerGluAlaPheCysHis 85
Db 2 TACCTGGAGATCAGCTGCGGTGACCGCTTGCGCGCAGCAGCGCGCTTCTGCCAC 61
QY 86 ArgAsnCysSerAlaAlaProGlnProGluProAlaAlaGlyLeuAlaSerTyrProGlu 105
Db 62 CGCACTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 121
QY 106 LeuArgLeuPheGlyGlyLeuLeuArgArgAlaHisCysLeuLeuArgCysGlyGly 125
Db 122 CTGGCGCTCTTCGGGGGCGCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 181
QY 126 LeuProAlaPheArgGlnSerGlnProSerArgGluValLeuAlaAspPheGlnArg 145
Db 182 CTGCCAGCCTTCGCGCAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 241

```

146 GluProTyrLysPheLeuGlnPheAlaTyrPheLysAlaAsnLeuProLysAlaIle 165  
242 GAGCCCTCAAGTCTCTGAGTTCGTTACTTCAAGGCAATAATCTCCCAAGGCATC 301  
QY 166 AlaAlaAlaHisThrPheLeuLeuLysHisProAspAspGluMetMetLysArgAsnMet 185  
Db 302 GCGCTGCTCACACCTTTCTACTGAAGCATCTCTGATGAGCAATGATGAAGAGGACATG 361  
QY 186 AlaTyrTyrLysSerLeuProGlyAlaGluAspTyrIleLysAspLeuGluThrLysSer 205  
Db 362 GCATATTATAAGACCTGCTGTCGCGGAGGACTACATTAAAGACCTGGAACCAAGTCA 421  
QY 206 TyrGluSerLeuPheIleArgAlaValArgAlaTyrAsnGlyGluAsnTrpArgThrSer 225  
Db 422 TATGAAGGCTGTTCATCCGAGCGTCCGCGCATACACGGTGAGAACCTGGGAGAACATCC 481  
QY 226 IleThrAspMetGluLeuAlaLeuProAspPhePheLysAlaPheTyrGluCysLeuAla 245  
Db 482 ATCAGACATGAGGCTGCGCTTCCGACTTCTTCAAGACCTTTACGAGTGTCTCGCA 541  
QY 246 AlaCysGluGlySerArgGluLeuLysAspPheLysAspPheTyrLysSerIleAlaAsp 265  
Db 542 GCCTGGCGGGTTCACAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAGAT 601  
QY 266 HisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsnLeuThrProValIle 285  
Db 602 CATTATGTAGAGTTCTGGAATGCAAAATACAGTGTCAAGAGAACCTCACCAGTTATA 661  
QY 286 GlyGlyTyrProValGluLysPheValAlaThrMetTyrHisTyrLeuGlnPheAlaTyr 305  
Db 662 GGAGGCTATCCGGTTGAGAAATTTGGTCTACCATGTATCATTTACGAGTTGGCTAT 721  
QY 306 TyrLysLeuAsnAspLeuLysAsnAlaAlaProCysAlaValSerTyrLeuLeuPheAsp 325  
Db 722 TATAAGTTGAACACCTGAAGATGAGGCCCTGTGSCAGTCAGCTATCTGCTTTTGTAT 781  
QY 326 GlnAsnAspLysValMetGlnGlnAsnLeuValTyrTyrGlnTyrHisArgAspThrTrp 345  
Db 782 CAGATGACAAAGTCTATGCAA-----GCAGAACCTTGG 814  
QY 346 GlyLeuSer 348  
|||:||||:  
815 GGTATTACC 823

RESULT 15  
BM550012  
LOCUS  
DEFINITION BM550012 1109 bp mRNA linear EST 20-FEB-2002  
5', mRNA sequence.  
ACCESSION BM550012  
VERSION BM550012.1 GI:18785873  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1109)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1920 row: i column: 17  
High quality sequence start: 20  
High quality sequence stop: 714.



## FEATURES

## Location/Qualifiers

1. .1109  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5443288"  
 /tissue\_type="astrocytoma grade IV, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 98"  
 /note="Organ: Brain; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."

BASE COUNT 277 a 331 c 282 g 219 t

## ORIGIN

## Alignment Scores:

Pred. No.: 3.1e-150 Length: 1109  
 Score: 1419.00 Matches: 269  
 Percent Similarity: 98.55% Conservative: 3  
 Best Local Similarity: 97.46% Mismatches: 1  
 Query Match: 66.62% Indels: 3  
 DB: 12 Gaps: 0

US-09-729-674-2 (1-401) x BM550012 (1-1109)

QY 84 CysHisArgAsnCyssSerAlaAlaProGlnProGluProAlaAlaGlyLeuAlaSerTyr 103  
 Db 33 TGCCACCGCACTGTAGCGCCGCGCGCGAGCGCGCGCGCTCGCCAGCTAT 92  
 QY 104 ProGluLeuArgLeuPheGlyGlyLeuLeuArgArgAlaHisCysLeuLysArgCysLys 123  
 Db 93 CCGAGCTCGCGCTCTCGCGGCGCTGCTCGCGCGCGCTGCTCAAGCGCTGCAAG 152  
 QY 124 GlnGlyLeuProAlaPheArgGlnSerGlnProSerArgGluValLeuAlaAspPheGln 143  
 Db 153 CAGGGCTGCCAGCTTCGCCAGTCCAGCCCGCGCGCGAGGCTGCTGGCGGACTTCCAG 212  
 QY 144 ArgArgGluProTyrLysPheLeuGlnPheAlaTyrPheLysAlaAsnLeuProLys 163  
 Db 213 CGCGCGGAGCCCTACAAAGTTCCTGAGTTCGTTACTTCAAGCAAAATATCTCCCAA 272  
 QY 164 AlaAlaAlaAlaHisThrPheLeuLeuLysHisProAspAspGluMetMetLysArg 183  
 Db 273 GCCATCGCCCTGCTCACACCTTCTACTGAGCATCCTGTAGCAAAATGATGAAGAGG 332  
 QY 184 AsnMetAlaTyrTyrLysSerLeuProGlyAlaGluAspTyrIleLysAspLeuGluThr 203  
 Db 333 AACATGGCATATTATAAGAGCTGCTGCTGCGGAGGACTACATTAAAGACCTGGAAACC 392  
 QY 204 LysSerTyrGluSerLeuPheLeuArgAlaValArgAlaTyrAsnGlyGluAsnTrpArg 223  
 Db 393 AAGTCATATGAAGCTGTTCATCCGAGCAGTCCGGGCATCAACGCTGAGAACCTGGAGA 452  
 QY 224 ThrSerIleThrAspMetGluLeuAlaLeuProAspPhePheLysAlaPheTyrGluCys 243  
 Db 453 ACATCCATCACAGACATGGAGCTGGCCCTTCCGACTTCTTCAAGCCCTTTACGAGTGT 512  
 QY 244 LeuAlaAlaCysGluGlySerArgGluIleLysAspPheLysAspPheTyrLeuSerIle 263  
 Db 513 CTCGCAGCCTGCCAGGGTTCACAGGAGATCAAGGACTTCAAGGANTTCTACCTTTCCATA 572  
 QY 264 AlaAspHisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsnLeuThrPro 283  
 Db 573 GCGAGATCATTTAGTAGAGTCTGGAAATGCAAAATACAGTGTGAGAGAACCTCACCCCA 632  
 QY 284 ValIleGlyGlyTyrProValGluLysPheValAlaThrMetTyrHisTyrLeuGlnPhe 303  
 Db 633 GTTATAGGAGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCTTACCTTGCAGTTT 692

QY 304 AlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAlaProCysAlaValSerTyrLeuLeu 323  
 Db 693 GCCTATTATTAAGTTGAACGACCTGAAGAATGCAGCCCTGTGCAGTCAGCTATCTGCTC 752  
 QY 324 PheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyrTyrGlnTyrHisArgAsp 343  
 Db 753 TTTGATCAGAAATGAACAGGTTCATGCAGCAGAACTGCTGTATTACCAGTACCACAGGGAA 812  
 QY 344 -ThrTrp-GlyLeuSerAspGluHisPhe-GlnProArgPro 356  
 Db 813 CACTTGGGGGCTCTCGGATGAGCACTCCACAGCCAGACCC 854

Search completed: January 30, 2004, 07:55:45  
 Job time : 2729 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 30, 2004, 03:39:47 ; Search time 357 Seconds  
(without alignments)  
3032.142 Million cell updates/sec

Title: US-09-729-674-2

Perfect score: 2130

Sequence: 1 MFPGRRGAALALLICVACA.....DDEGEVVEYVDDLLELETS 401

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q/cpn2\_1/USPTO.spool/US09729674/runat\_29012004\_095313\_12879/app\_query.fasta\_1.583  
-DB=N\_Geneseq\_19Jun03 -Qfmt=fastp -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09729674 @CN1\_1\_0 @runat\_29012004\_095313\_12879 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DSV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2128	99.9	3871	20	AAK60801	Human secreted pro
2	2128	99.9	3871	22	AAS9207	Human cDNA encodin
3	2128	99.9	3871	24	ABA90876	Human polynucleoti
4	2121	99.6	1961	22	AAH14525	Human cDNA sequenc
5	1247	58.5	1586	21	AAH31138	Human colon cancer
6	1155	54.2	813	22	AAH06808	Human cDNA clone (
7	1088.5	51.1	2347	24	ABH67662	Cesophagus cancer
8	717	33.7	408	21	AAH30585	Human colon cancer
9	574	26.9	2242	24	ABZ11346	Human polynucleoti
10	572.5	26.9	2993	22	AAH16288	Human cDNA sequenc
11	571	26.8	2211	24	AAH17573	DNA encoding novel
12	571	26.8	2524	22	AAH14513	Human cDNA sequenc
13	571	26.8	2563	22	AAH93800	Human cDNA encodin
14	571	26.8	2585	22	AAH14589	Human cDNA sequenc
15	568	26.7	1309	24	ABZ11344	Human polynucleoti
16	567	26.6	2600	21	AAH64725	Human tumour suppr
17	562.5	26.4	2577	22	AAH08489	Human secreted pro
18	549	25.8	2152	24	ABZ11345	Human polynucleoti
19	543.5	25.5	2615	22	AAH08507	Human secreted pro
20	539	25.3	2829	21	AAH64724	Human tumour suppr
21	538	25.3	2322	21	AAH64727	Mouse tumour suppr
22	538	25.3	2416	21	AAH64726	Mouse tumour suppr
23	527	24.7	2753	24	ABX94910	Human novel polynu
24	527	24.7	2753	24	ABX94913	Human novel polynu
25	473.5	22.2	2127	24	AAH17572	DNA encoding novel
26	456	21.4	2665	25	ACA03965	cDNA downregulated
27	382.5	18.0	2789	22	AAH14459	Human cDNA sequenc
28	376.5	17.7	800	22	AAH06448	Human cDNA clone (
29	340.5	15.5	1025	24	ABQ27312	Oligonucleotide fo
30	330.5	15.5	1025	24	ABQ27313	Oligonucleotide fo
31	321.5	15.1	2305	25	ABX70413	DNA encoding human
32	306.5	14.4	1026	24	ABQ26350	Oligonucleotide fo
33	306.5	14.4	1026	24	ABQ26351	Oligonucleotide fo
34	302	14.2	550	25	ACA57060	Human adipocyte se
35	283.5	13.3	774	22	AAH06885	Human cDNA clone (
36	273.5	12.8	652	22	AAH93398	Primer specific fo
37	273.5	12.8	703	22	AAH06792	Human cDNA clone (
38	268	12.6	1026	24	ABQ26352	Oligonucleotide fo
39	268	12.6	1026	24	ABQ26353	Oligonucleotide fo
40	267	12.5	1025	24	ABQ27310	Oligonucleotide fo
41	267	12.5	1025	24	ABQ27311	Oligonucleotide fo
42	285.5	12.5	738	22	AAH06612	Human cDNA clone (
43	252	11.8	963	23	AAH80465	DNA encoding novel
44	214.5	10.1	682	22	AAH96734	Human neuroblastom
45	200.5	9.4	349	21	AAH18376	Lung cancer associ

# ALIGNMENTS

RESULT 1

AAK60801

ID AAK60801 standard; DNA; 3871 BP.

XX AAK60801;

XX 09-AUG-1999 (first entry)

DT Human secreted protein encoding DNA (clone bd306-7).

DE Secreted protein; kidney; lung; brain; blood; testis; bone marrow;  
KW nutritional activity; cytokine; cell proliferation; immune stimulation;  
KW hematopoiesis regulation; tissue growth; thrombolytic; gene therapy;  
KW anti-inflammatory; tumour invasion; ss.

XX Homo sapiens.

OS

XX





QY 341 HisArgAspThrTrpGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGln 360  
 Db 1083 CACAGGACACTTGGGGCTCTCGGATGAGCACTTCAGCCACAGCTGAGCAGTTGAG 1142  
 QY 361 PhePheAsnValThrThrLeuGlnLysGluLeuTyAspPheAlaLysGluAsnIleMet 380  
 Db 1143 TTCTTTAATGTGACCACTCCAGAGGAGCTGTATGACTTTCTAAGGAATAATATG 1202  
 QY 381 AspAspAspGluGlyGluValValGluTyValAspAspLeuLeuGluGluThr 400  
 Db 1203 GATCATGATGAGGAGAGAGTGTGTGAATATGTGGATGACCTCTTGAACCTGGAGGAGACC 1262  
 QY 401 Ser 401  
 Db 1263 AGC 1265  
 RESULT 3  
 ABA90876  
 ID ABA90876 standard; cDNA; 3871 BP.  
 XX ABA90876;  
 AC ABA90876;  
 DT 14-FEB-2002 (first entry)  
 XX Human polynucleotide SEQ ID NO 1.  
 DE Human; clone bd306-7; ATCC number 98599; gene therapy;  
 KW immune disorder; bacterial infection; fungal infection; cancer; tumour;  
 KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;  
 KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;  
 KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnerability;  
 KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;  
 KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;  
 KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.  
 XX Homo sapiens.  
 OS  
 XX US2001039335-A1.  
 PN  
 XX 08-NOV-2001.  
 PD  
 XX 04-DEC-2000; 2000US-0729674.  
 PF  
 XX 10-APR-1997; 97US-126425P.  
 PR 04-DEC-1997; 97US-067454P.  
 PR 20-DEC-1997; 97US-068379P.  
 PR 02-JAN-1998; 98US-070346P.  
 PR 07-JAN-1998; 98US-070643P.  
 PR 08-JAN-1998; 98US-070755P.  
 PR 13-JAN-1998; 98US-071304P.  
 PR 22-JAN-1998; 98US-072134P.  
 PR 30-JAN-1998; 98US-073095P.  
 PR 18-FEB-1998; 98US-075038P.  
 PR 30-MAR-2000; 2000US-0539330.  
 PR 23-NOV-1998; 98US-0197886.  
 XX (JACO/) JACOBS K.  
 PA (MCCO/) MCCOY J M.  
 PA (LAVA/) LAVALLIE E R.  
 PA (COLL/) COLLINS-RACIE L A.  
 PA (EVAN/) EVANS C.  
 PA (MERB/) MERBERG D.  
 PA (TREA/) TREACY M.  
 PA (AGOS/) AGOSTINO M J.  
 PA (STEI/) STEININGER R J.  
 PA (SPAU/) SPAULDING V.  
 PA (WONG/) WONG G G.  
 PA (CLAR/) CLARK H.  
 PA (FECH/) FECHTEL K.  
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

PI Wong GG, Clark H, Fechtel K;  
 XX WPI: 2002-040725/05.  
 DR P-PSDB; ABB55698.  
 XX  
 PT New secreted proteins and encoding polynucleotides, useful in gene  
 PT therapies, particularly for preventing or treating autoimmune  
 PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,  
 PT stroke or inflammations  
 XX  
 PS Claim 1; Page 175-177; 349pp; English.  
 XX  
 CC The invention relates to isolated polynucleotides (ABA90876-ABA90968 and  
 CC ABA90980) and encoded proteins (ABB55698-ABB55800), especially  
 CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and  
 CC proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in  
 CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1  
 CC are deposited with the American Type Culture Collection (ATCC) with  
 CC accession number 98599. The polynucleotides and encoded polypeptides have  
 CC cytostatic, anti-inflammatory, immunomodulator, vulnerary,  
 CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic  
 CC and anti-inflammatory activity and acting as cytokine modulators,  
 CC haematopoiesis regulators, tissue growth modulators and/or cadherin  
 CC suppressors. The polypeptides and polynucleotides are useful in gene  
 CC therapies, particularly for preventing, treating or ameliorating any of  
 CC the following diseases: immune deficiency and disorders; e.g. bacterial  
 CC or fungal infections, autoimmune disorders, cancer, systemic lupus  
 CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell  
 CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
 CC osteoarthritis; central and peripheral nervous system diseases and  
 CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
 CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;  
 CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis  
 CC or systemic inflammatory response syndrome, ischaemia-reperfusion  
 CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or  
 CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus  
 CC foliaceus.  
 XX  
 SQ Sequence 3871 BP; 965 A; 913 C; 1033 G; 958 T; 2 other;  
 Alignment Scores:  
 Pred. No.: 7.01e-201 Length: 3871  
 Score: 2128.00 Matches: 401  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.91% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-729-674-2 (1-401) x ABA90876 (1-3871)  
 QY 1 MetGluProGlyArgArgGlyAlaAlaAlaLeuAlaLeuLeuLeuValAlaCysAla 20  
 Db 63 ATGGAGCGGGGCGCGCGGGGGCGCGCGCTGCTAGCGCTGTGTGCTGGCTCGCG 122  
 QY 21 LeuArgAlaGlyArgAlaGlnTyrGluArgTyrSerPheArgSerPheProArgAspGlu 40  
 Db 123 CTGGCGCGGGGCGCGCGCGCCCAATACGACGCTTCCGAGCTTCCACGGGACGAG 182  
 QY 41 LeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGlyGluHisTrp 60  
 Db 183 CTGATGCGGCTCGAGTCGGGCTTACCGCACGCGCTGGACGACGCGGCGGAGCACTGG 242  
 QY 61 AlaGluSer\*\*\*GlyTyrLeuGluIleSerLeuArgLeuHisArgLeuLeuArgAspSer 80  
 Db 243 GCCGAGAGCSTKGGCTACCTGGAGATCAGCTCGCGCTGCACCGCTTGTCTGCGGACAGC 302  
 QY 81 GluAlaPheCysHisArgAsnCysSerAlaAlaProGlnProGluProAlaAlaGlyLeu 100  
 Db 303 GAGGCTTCTGCCACCGCAACTGCGAGCGCGCGCGCGAGCCCGCGCGCGCCCTC 362  
 QY 101 AlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArgArgAlaHisCysLeuLys 120  
 Db 363 GCCAGCTATCCGAGCTGCGCTTTCGGGGGCGCTTCTGGCGCGCGGACACTGCTCAAG 422

```
QY 121 ArgCysLysGlnGlyLeuProAlaIleAlaHisThrPheLeuLeuLysAspGluValLeuAla 140
Db 122 CGTGTCAAGCAGGCGCTGCGAGCTTCCGCCAGTCCAGCCAGCGCGAGGTGTGCGG 482
QY 141 AspPheGlnArgArgGluProTyrLysPheLeuGlnPheAlaTyrPheLysAlaAsnAsn 160
Db 142 GACTTCCAGCGCGCGAGCCCTACAGTTCTCGAGTTCTGCTTACTTCAAGGCAATATAT 542
QY 161 LeuProLysAlaIleAlaAlaHisThrPheLeuLeuLysHisProAspAspGluMet 180
Db 162 CTCCCAAGACCATCCCGCTGCTCAGCTTCTTCTACTGAACATCCCTGATCAGCAATG 602
QY 181 MetLysArgAsnMetAlaTyrTyrLysSerLeuProGlyValAlaGluAspTyrLysAsp 200
Db 182 ATGAAGAGGAACATGSCATATATATAGAGCTTCCCTGCTGCGAGGACTACATTAAGAC 662
QY 201 LeuGluThrLysSerTyrGluSerLeuPheLeuAlaValArgAlaTyrAsnGlyGlu 220
Db 202 CTGGAACCAAGTCATATGAAGCCTGTTTCATCCGAGCAGTCCGCGCATACAACTGAG 722
QY 221 AsnTyrArgThrSerLeuThrAspMetGluLeuAlaLeuProAspPhePheLysAlaPhe 240
Db 222 ACTTGAGAACATCCATCAGACATGAGCTGGGCTTCCCGATCTTCTTCAAGCCCTT 782
QY 241 TyrGluCysLeuAlaAlaCysGluGlySerArgGluIleLysAspPheLysAspPheTyr 260
Db 242 TAGAGTGTCTCGACCTGCGAGGCTTCCAGGAGATCAAGGACTTCAAGGATTTCTAC 842
QY 261 LeuSerIleAlaAspHisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsn 280
Db 262 CTTTCATAGCAGATCATATATGAAGTCTGGAATGCAAAATACAGTGTGAAGAGAAC 902
QY 281 LeuThrProValIleGlyTyrProValGluLysPheValAlaThrMetTyrHisTyr 300
Db 282 CTCACCCAGTATAGGAGCTATCCGTTGAGAAATTTGCGCTACCATGTATCATTTAC 962
QY 301 LeuGlnPheAlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAlaProCysAlaValSer 320
Db 302 TTGCGATTTGCTTATTAAGTTGAACGACCTGAAGATGTCAGCCCTGTCAGTCAGC 1022
QY 321 TyrLeuLeuPheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyrGlnTyr 340
Db 322 TATCTGCTTTTATCAGAAATGACAGGTCTGCGAGAGAACTCGTGTATTACCACTAC 1082
QY 341 HisArgAspThrTrpGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGln 360
Db 342 CACAGGACACTTGGGCTCTCGATGAGCATTCCAGCCAGACCTGAAGCAGCTTCAG 1142
QY 361 PhePheAsnValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGluAsnIleMet 380
Db 362 TTCTTTAATGTGACCACACTCCAGAGGAGCTGTATGCTTTCCTAAGGAAATATATG 1202
QY 381 AspAspAspGluGlyGluValValGluTyrValAspAspLeuLeuGluGluThr 400
Db 382 GATGATGATGAGGAGAGCTTGTGGAATATGTGATGATGACCTTCTGGAAGTGGAGGAC 1262
QY 401 Ser 401
Db 1263 AGC 1265

RESULT 4
ID AAH14525 standard; cDNA; 1961 BP.
XX AAH14525;
XX AC
XX ATGAGAGCGGGCGCGCGGGCGCGGGCGCGGGCGCGGGCGCGGGCGGTGGTGGCGTCCGCG 99
DT 26-JUN-2001 (first entry)
XX 21 LeuArgAlaGlyArgAlaGlnTyrGluArgTyrSerPheArgSerPheProArgAspGlu 40
XX 100 CTGCGCGCGCGCGCGCGCGCCCAATACAGCTTCCGAGCTTCCGCGGACGAG 159
DE Human cDNA sequence SEQ ID NO:12066.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX 41 LeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGlyGluHisTrp 60
```

```
OS Homo sapiens.
XX EPI074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300053.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX Claim 8; SEQ ID 12066; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
SQ Sequence 1961 BP; 417 A; 579 C; 481 G; 484 T; 0 other;

Alignment Scores:
Pred No.: 1-35e-200 Length: 1961
Score: 2121.00 Matches: 399
Percent Similarity: 99.50% Conservative: 0
Best Local Similarity: 99.50% Mismatches: 2
Query Match: 99.58% Indels: 0
DB: 22 Gaps: 0

US-09-729-674-2 (1-401) x AAH14525 (1-1961)
QY 1 MetGluProGlyArgArgGlyAlaAlaLeuLeuAlaLeuLeuLeuLeuLeuLeuLeuAla 20
Db 40 ATGAGAGCGGGCGCGCGGGCGCGGGCGCGGGCGCGGGCGGTGGTGGCGTCCGCG 99
QY 21 LeuArgAlaGlyArgAlaGlnTyrGluArgTyrSerPheArgSerPheProArgAspGlu 40
Db 100 CTGCGCGCGCGCGCGCGCCCAATACAGCTTCCGAGCTTCCGCGGACGAG 159
QY 41 LeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGlyGluHisTrp 60
```

Db 160 CTGATGCCGCTCGAGTCCGCTACCGGCACGCGCTGGGCAAGTACAGCGCGGAGCACTGG 219  
Qy 61 AlaGluSer\*\*GlyTyrLeuGluLeuHisArgLeuLeuArgAspSer 80  
Db 220 GCGAGAGCGTGGGCTACCTGGAGATCAGCTCGGCTGCACCGCTTGTCTGCGGACAGC 279  
Qy 81 GluAlaPheCysHisArgAsnCysSerAlaAlaProGlnProGluProAlaAlaGlyLeu 100  
Db 280 GAGGCCCTTCGCCACCGCACTGAGCGCGCGCGCGAGCGCGCGCGCGCGCTC 339  
Qy 101 AlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArgAlaHisCysLeuLys 120  
Db 340 GCCAGCTATCCGAGCTCGGCTCTTCGGGGGCTCTGCGCGCGCGCACTGCCTCAAG 399  
Qy 121 ArgCysLysGlnGlyLeuProAlaPheArgGlnSerGlnProSerArgGluValLeuAla 140  
Db 400 CGTGCAGAGCGGCGCTCCAGCGCTTCGCGCAGTCCCGAGCGCGAGCGGCTGGCG 459  
Qy 141 AspPheGlnArgArgGluProTyrLysPheLeuGlnPheAlaTyrPheLysAlaAsnAsn 160  
Db 460 GACTTCCAGCGCGCGAGCGCTACAGTTCTCGAGTTCGCTTACTTCAAGGCAATAT 519  
Qy 161 LeuProLysAlaLeuAlaAlaHisThrPheLeuLeuLysHisProAspAspGluMet 180  
Db 520 CTCGCCAAAGCATCGCGCTGTCTCACACCTTCTACTGAAGCATCTGATGACGAAATG 579  
Qy 181 MetLysArgAsnMetAlaTyrTyrLysSerLeuProGlyAlaGluAspTyrLysAsp 200  
Db 580 ATGAAGAGGACATGGCATATTAAAGCGCTGCTGTCGCGAGGACTACATTAAGAC 639  
Qy 201 LeuGluThrLysSerTyrGluSerLeuPheLeuAlaValArgAlaTyrAsnGlyGlu 220  
Db 640 CTGGAACCAAGTCATATGAAGCGCTGTTCATCCGAGCAGTCCGGGCATACACAGGTGAG 699  
Qy 221 AsnTrpArgThrSerIleThrAspMetGluLeuAlaLeuProAspPhePheLysAlaPhe 240  
Db 700 AACTGGAGACATCCATCAGACATGGAGCTGGCGCTTCCGACCTTCTTCAAGCGCTT 759  
Qy 241 TyrGluCysLeuAlaAlaCysGluGlySerArgGluLysAspPheLysAspPheTyr 260  
Db 760 TACGAGTGTCTCGCAGCGCTCGAGGCTTCAGGGAGATCAAGGACTTCAAGGATTTCTAC 819  
Qy 261 LeuSerIleAlaAspHisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsn 280  
Db 820 CTTTCCATAGCAGATCATATTATGAGAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAAC 879  
Qy 281 LeuThrProValIleGlyGlyTyrProValGluLysPheValAlaThrMetTyrHisTyr 300  
Db 880 CTCACCCAGTTATAGAGGCTATCCGGTTGAGAAATTTGGCTACCATGATCATTTAC 939  
Qy 301 LeuGlnPheAlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAlaProCysAlaValSer 320  
Db 940 TTGCAGTTTGCCTATTATAGTTGAACGACCTGAAGAATCGAGCCCGCTGTGCAGTCAGC 999  
Qy 321 TyrLeuLeuPheAspGlnAsnAspLysValMetClnGlnAsnLeuValTyrGlnTyr 340  
Db 1000 TATCTGCTTTGTATCAGATGACAGGTATGACAGAACCTTGGTGTATACCAAGTAC 1059  
Qy 341 HisArgAspThrTrpGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGln 360  
Db 1060 CACAGGACACGTGGGCGCTCTCAGATGAGCCTTCCAGCCCGAGCGCTGAAGCAGTTTCAG 1119  
Qy 361 PhePheAsnValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGluAsnIleMet 380  
Db 1120 TTCTTTTATGTGACACACCTCCAGAGGAGCTGTATGACTTTGCTAAGGAAATATAATG 1179  
Qy 381 AspAspAspGluGlyGluValValGluTyrValAspAspLeuLeuGluLeuGluThr 400  
Db 1180 GATCATGATGAGGAGAGAGTGTGGATATGTGGATGACCTCTTGGNACTGGAGGAGACC 1239  
Qy 401 Ser 401  
Db 1240 AGC 1242

## RESULT 5

AAH31138  
ID AAH31138 standard; cDNA; 1586 BP.  
XX  
XX AAH31138;  
AC  
XX 27-JUL-2001 (first entry)  
DT Human colon cancer cell line Km12L4-A cDNA library derived seq#1072.  
DE Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;  
KW detection; colon cancer cell line Km12L4-A; ss.  
KW  
XX Homo sapiens.  
OS  
XX WO200018916-A2.  
XX  
XX 06-APR-2000.  
XX  
XX 23-SEP-1999; 99WO-US22226.  
XX  
XX 28-SEP-1998; 98US-0102161.  
PR 28-SEP-1998; 98US-0102180.  
PR 29-SEP-1998; 98US-0102380.  
PR 08-OCT-1998; 98US-0103815.  
PR 27-OCT-1998; 98US-0105877.  
XX  
XX (CHIR) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
XX  
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
PI Leskowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
XX  
XX WPI; 2000-299155/25.  
XX  
XX Polynucleotide library comprising 1079 defined sequences, useful in  
PT the form of an array to detect cancer or susceptibility to cancer -  
PT  
XX  
XX Claim 1; Page 498-499; 502pp; English.  
XX  
XX The present invention describes a library of polynucleotides comprising  
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described  
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to  
CC one of the 1079 sequences; (2) a recombinant host cell containing (I);  
CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that  
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method  
CC of detecting differentially expressed genes correlated with a cancerous  
CC state of a mammalian cell comprising detecting a gene product encoded by  
CC 65 of the 1079 sequences given in the specification. The polynucleotides  
CC are used to monitor patients having (or susceptible) to cancer to detect  
CC potentially malignant events at a molecular level before they are  
CC detectable at a gross morphological level. The polynucleotides are also  
CC useful for monitoring the efficacy of various therapies and preventive  
CC interventions. Polynucleotide probes based on the disclosed sequences  
CC are useful for chromosome mapping and detection of transcription levels.  
CC The 1079 polynucleotide sequences were derived from a human colon cancer  
CC cell line Km12L4-A cDNA library.  
XX  
XX Sequence 1586 BP; 381 A; 406 C; 340 G; 453 T; 6 other;  
SQ  
Alignment Scores:  
Pred. NO.: 6.39e-114 Length: 1586  
Score: 1247.00 Matches: 233  
Percent Similarity: 99.15% Conservative: 0  
Best Local Similarity: 99.15% Mismatches: 2  
Query Match: 58.54% Indels: 0  
DB: 21 Gaps: 0  
US-09-729-674-2 (1-401) X AAH31138 (1-1586)



```
QY 167 AlaAlaHisThrPheLeuLeuLysHisProAspAspGluMetMetLeuArgAsnMetAla 186
DB 3 GCTGCTCACACCTTTCTACTAGCATCTCTGATGACGAAATGATGAAGAGAAACATGGCA 62
QY 187 TyrTyrLysSerLeuProGlyAlaGluAspTyrIleLysAspLeuGluThrLysSerTyr 206
DB 63 TATTATAAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 122
QY 207 GluSerLeuPheIleArgAlaValArgAlaTyrAsnGlyGluAsnTrpArgThrSerIle 226
DB 123 GAAAGCCTTCTCATCGAGCAGTGGGGCATACAACGCTGAGAACTGGAGAACATCCATC 182
QY 227 ThrAspMetGluLeuAlaLeuProAspPheLysAlaPheTyrGluCysLeuAla 246
DB 183 ACAGACATGGAGTGGCTTCCGCTTCTTCAAGGCTTTTACGAGTCTGCGAGCC 242
QY 247 CysGluGlySerArgGluIleLysAspPheLysAspPheTyrLeuSerIleAlaAspHis 266
DB 243 TCCRRGGGTTCCAGGAGATCAAGCATTTCAAGGATTTCTACTTTCATAGCAGATCAT 302
QY 267 TyrValGluValLeuGluCysLysIleGlnCysGluGluAsnLeuThrProValIleGly 286
DB 303 TATGTAGAGTTCGTGAATGCAAAATACAGTCTGAAGAGAACTCACCCAGTTATAGGA 362
QY 287 GlyTyrProValGluLysPheValAlaThrMetTyrHisTyrLeuGlnPheAlaTyrTyr 306
DB 363 GACTATCCGGTGGAGAAATTTGTGCTACCATCATATCTACTTGCAGTTTGCCTATTAT 422
QY 307 LysLeuAsnAspLeuLysAsnAlaProCysAlaValSerTyrLeuLeuPheAspGln 326
DB 423 AAGTTGAACGACCTGAAGAAATGACGCCCTCTGCGAGTCAGTACTCTCTTGTATCAG 482
QY 327 AsnAspLysValMetGlnGlnAsnLeuValTyrTyrGlnTyrHisArgAspThrTrpGly 346
DB 483 AATGCAAGGTCATGACGAGAACCTGGTGATTTACCATGACACAGGACACATGGGGC 542
QY 347 LeuSerAspGluHisPheGlnProArgProGluAlaValGlnPheAsnValThrThr 366
DB 543 CTCTCKGATGACACTTCCAGCCAGACCTGAAGCAGTTTCAGTTCTTTAATGTGACCACA 602
QY 367 LeuGlnLysGluLeuTyrAspPheAlaLysGluAsnIleMetAspAspGluGlyGlu 386
DB 603 CTTCCAGAGGCTGTATGACTTTCTTAAAGGAAATATAATGATGATGATGATGAGGAGAA 662
QY 387 ValValGluTyrValAspAspLeuLeuGluGluLeuGluThrSer 401
DB 663 GTTGTGGATATGTGATGACCTCTTGGAACTGGAGGAGACCAGC 707

RESULT 6
AAH06808
ID AAH06808 standard; cDNA; 818 BP.
XX
AC AAH06808;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:3643.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
```

```
XX (HELI-) HELIX RES INST.
PA Ota T, Isozaki T, Nishikawa T, Hayaehi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 3643; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 818 BP; 159 A; 271 C; 236 G; 148 T; 4 other;

Alignment Scores:
Pred. No.: 3.5e-105 Length: 818
Score: 1155.00 Matches: 238
Percent Similarity: 91.57% Conservative: 1
Best Local Similarity: 91.19% Mismatches: 15
Query Match: 54.23% Indels: 8
DB: 22 Gaps: 1

US-09-729-674-2 (1-401) x AAH06808 (1-818)
QY 1 MetGluProGlyArgArgGlyAlaAlaAlaLeuLeuAlaLeuLeuLeuLeuLeuLeuLeuLeu 210
DB 40 ATGGAGCGGGGGCGCGGGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGG 99
QY 21 LeuArgAlaGlyArgAlaGlnTyrGluA:GlyTyrSerPheArgSerPheProArgAspGlu 40
DB 100 CTGCGCGCGGGGGCGCGGGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGG 159
QY 41 LeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGlyGluHisTtp 60
DB 160 CTGATGCCGCTCGAGTCGGCTTACCGGCGCGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGG 219
QY 61 AlaGluSer***GlyTyrLeuGluIleSerLeuArgLeuHisArgLeuLeuArgAspSer 80
DB 220 GCCGAGAGCGTGGGCTACCTGGAGATCAGCTCGGGCTGCACCGCTTGTGCGGACAGC 279
QY 81 GluAlaPheCysHisArgAsnCysSerAlaAlaProGlnProGluProAlaAlaGlyLeu 100
DB 280 GAGGCTTCTGCCACCGCAACTGCAGCGCGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGG 339
QY 101 AlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArgArgAlaHisCysLeuLys 120
DB 340 GCCAGCTATCCGAGCTGCGCTTTCGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 399
```



Qy	44	LeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGlyGluHisTirPAlaGluSer	63
Db	117	CTGCCCGGGGTACGGCAGCGCTCTGGAGCATACGAGGAGAGAGCTGCGCGGAGAGC	176
Qy	64	**GlyTyrLeuGluIleSerLeuArgLeuHisArgLeuLeuArgAspSerGluAlaPhe	83
Db	177	GCGCGCTACCTGGAGCGCGCGCTGCGGTGTGACCGGTCTCTGGCGCAGCAGGAGCCTTC	236
Qy	84	CysHisArgAsnCysSerAla-----AlaProGlnProGluProAlaAlaGlyLeu	100
Db	237	TGCCACGCAACTGACGCGCGCCCGCGCCGCGCCAGCCGATCCCGACGCGCGCGCGC	296
Qy	101	AlaSer-----TyrProGluLeuArgLeuPheGlyGlyLeuLeuArgArgAlaHisCys	118
Db	297	GCACAGAGTGGCGCTGCGAGCTGGCGCTCTTCGGCCGCGCTCTGGAGCAGCGCCCTGC	356
Qy	119	LeuLysArgCysLysGlnGlyLeuProAlaPheArgGlnSerGlnProSerArgGluVal	138
Db	357	CTGGCGCGCTCAAGCGGACGCTGCCGCCCTTCCAGGTGCCCTACCCGCGCGCGCAGCTG	416
Qy	139	LeuAlaAspPheGlnArgArgGluProTyrLysPheLeuGlnPheAlaTyrPheLysAla	158
Db	417	CTCGGTGACTTCCAGAGCGCGCTGCCCTACCAGTACTGCTACTAGCGCGCTGTTCAAGGCT	476
Qy	159	AsnAsnLeuProLysAlaIleAlaAlaHisThrPheLeuLeuLysHisProAspAsp	178
Db	477	AACCGCTGGAGAGCGGTGGCGCGCCTACACCTTCTCCAGAGGAAACCCGAGCAC	536
Qy	179	GluMetMetLysArgAsnMetalTyrTyrLysSer---LeuProGlyAlaGluAspTyr	197
Db	537	GAGCTACCGCCCAAGTATCTCAACTACTATCAGGGGATGCTGGACGCTGCGCAGAGTCC	596
Qy	198	IleLysAspLeuGluThrLysSerTyrGluSerLeuPheIleArgAlaValArgAlaTyr	217
Db	597	CTCAGGACCTAGAGGCCACCCCTACGAGGCGGTGTTCTCCGGGTGTGGAGCTCTAC	656
Qy	218	AsnGlyGluAsnTrpArgThrSerIleThrAspMetGluLeuAlaLeuProAspPhePhe	237
Db	657	AACAGCGGGATTTCCGACGACGACGAGGACATGAGCGGGGCTTGTCCAGAGTACCTG	716
Qy	238	LysAlaPheTyrGluCysLeuAlaAlaCysGluGlySerArgGluIleLysAspPheLys	257
Db	717	GCACTCTTGGCCGGTGCTGGCGGCTGTGAAGGGGCCCATGACGAGGTGACTTCAAG	776
Qy	258	AspPheTyrLeuSerIleAlaAspHisTyrValGluValLeuGluCysLysIleGlnCys	277
Db	777	GACTTACCGGCCATAGCAGATCTCTTGCAGAGTCCCTGTCAGTGCAGGTGGACTGT	836
Qy	278	GluGlnAsnLeuThrProValIleGlyGlyTyrProValGluLysPheValAlaThrMet	297
Db	837	GAGCCCAATTGACCCCAATGTGGGTGGCTACTCTGTGGCAACAAGTTCGTGCCACCATTG	896
Qy	298	TyrHisTyrLeuGlnPheAlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAlaProCys	317
Db	897	TACCACCTACCTGAGTTTGGCTACTATAAGTTGAATGATGTGCCCGCAGGTGCCCGCAGC	956
Qy	318	AlaValSerTyrLeuLeuPheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyr	337
Db	957	GCCGCCAGCTACATGCTCTTCGACCCCAAGACAGCGCTCATGACAGCAACCTGCTGTAT	1016
Qy	338	TyrGlnTyrHisArgAspThrTrpGlyLeuSerAspGluHisPheGlnProArgProGlu	357
Db	1017	TACCGTTCCACCGGCTCGCTGGGGCTGTGAAGAGGAGGAGTCTCCAGCCCGGGAGGAG	1076
Qy	358	AlaValGlnPhePheAsnValThrLeuGlnLysGluLeuTyrAspPheAlaLysGlu	377
Db	1077	GCCATGCTCTACCAACAACAGACCGCGAGTGTGGGAGTGTCTGGAGTTCACCCACATG	1136
Qy	378	AsnIleMetAspAspAspGluGlyValValGluTyrValAspAspLeuLeuGluLeu	397
Db	1137	TACCTCAGTCAGATGATGAG-----ATGGAGCTG	1166
Qy	398	GluGluThr	400







Db 115 -----TTCCGCGAGGGGACCGCAGCTTACGGCGGGGGAGCTGGCCCGGGGTGGTCTCTG 168  
QY 66 TyrLeuGluIleSerLeuArgLeuHisArgLeuLeuArgAspSerGluAlaPheCysHis 85  
Db 169 AGCATGGAACGGGCGCTGGCTCCCGGGGAGCGCTCCGGCGCTTCGGCTGGCTGGCGGC 228  
QY 86 ArgAsnCysSerAla-----AlaProGlnProGlu 95  
Db 229 ACCCAGTGTGGCGCGCAGCTTCCCGTGGAGCTGGAGCCCGCAGCTGGTCCCGCCGCGGCC 288  
QY 96 ProAlaAlaGlyLeuAlaSerTyrProGluLeuArgLeuPheGlyGlyLeuArgArg 115  
Db 289 CAGGCTCGGGCGCGCGCTCCCGGACCTTCTCGGGGCTTCTGGGTGCGC 348  
QY 116 AlaHisCysLeuArgCysGlnGlyLeuProAlaPheArgGlnSerGlnProSer 135  
Db 349 GCTGCTGCTGCGCGCTGCG-----CTCGGGCGCGCGCGCCACCTCGCTC-----AGC 399  
QY 136 ArgGluValLeuAlaAspPheGlnArgGluProTyrLysPheLeuGlnPheAlaTyr 155  
Db 400 GAAGAGATGGAGTGGAGTTCGCAAGCGGAGCCCTCAACTACTGAGTGGCTGCTTAC 459  
QY 156 PheLysAlaAsnAsnLeuProLysAlaIleAlaAlaHisThrPheLeuLeuLysHis 175  
Db 460 TTCAAGATCAACAAGTTGAGAAAGCTGTGTGCGAGCACACACCTTCTGCTGGCAAT 519  
QY 176 ProAspAspGluMetLysArgAsnMetAlaTyrTyrLysSerLeuProGlyAlaGlu 195  
Db 520 CTTGAGCATGGAATGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 579  
QY 196 AspTyr---IleLysAspLeuGluThrLysSerTyrGluSerLeuPheIleArgAlaVal 214  
Db 580 GAGCGCGACTTCAGGATCTTGAGACTCAACCCCATATCGAAGATTTTCGACTGGAGTG 639  
QY 215 ArgAlaTyrAsnGlyGluAsnTrpArgThrSerIleThrAspMetGluLeuAlaLeuPro 234  
Db 640 CGACTCTACTAGAGAAACAGCACAGGAGAGTGTGCCCCACCTAGAGCGCGCGCTGCAA 699  
QY 235 AspPheLeuAlaPheTyrGluCysLeuAlaAlaCysGluGlySerArgGluLeuLys 254  
Db 700 GAATACTTTGTGGCTATAGAGTGGCGCTGCGCTGCGAGGCGCTTACTGACTACGAT 759  
QY 255 AspPheLys-----AspPheTyrLeuSerIleAlaAspHisTyrVal 268  
Db 760 GGTACAACTACTGAGTACAAAGCTGACCTTCCAGGCCATCACAGATCATATACATC 819  
QY 269 GluValLeuGluCysLysIleGlnCysGluGluAsnLeu---ThrProValIleGly 287  
Db 820 CAGGCTCTCACTGTAAGCAGAACTGTGTCAGGAGCTGTCTCCACCCCAAGTCGAGAG 879  
QY 288 TyrProValGluLysPheValAlaThrMetTyrHisTyrLeuGlnPheAlaTyrTyrLys 307  
Db 880 AAGCCCTTTGAGAGCTTCTCCCATCGCATATATATCTGCACTTTGCCCTACTATAC 939  
QY 308 LeuAsnAspLeuLysAsnAlaProCysAlaValSerTyrLeuLeuPheAspGlnAsn 327  
Db 940 ATTGGGAATTATACAGAGCTGTGAAATGTGCAAGACCTATCTTCTTCTTCCCAAT 999  
QY 328 AspLysValMetGluGlnAsnLeuValTyrTyrGlnTyrHisArgAspThrTrpGlyLeu 347  
Db 1000 GACGAGGTGATGAACCAAAATTTGGCTTATATGAGCAGTATG-----CTT 1044  
QY 348 SerAspGluHisPheGln-----ProArgProGluAlaValGlnPhePheAsnVal 364  
Db 1045 CGAAGAAGAACACACAGATCCATCGGCCCGCTGAGAGTGGCCAGAGGAGTACGACAGCGA 1104  
QY 365 ThrThrLeuGlyGluLeuTyrAspPheAlaLysGlu-----AsnIleMet 380  
Db 1105 AGCCTACTGGAAGAAAGACTGTTCTTCTCGCTTATGATGTTTTTGGAAATTCCTTTGTG 1164  
QY 381 AspAspAspGluGlyGluValValGluTyrValAspAspLeuLeuGluLeuGluThr 400  
Db 1165 GATCCGATTCATGGACTCCAGAGAGTGTATCCCAAGAGATTCAGAGAGAACAGAG 1224

QY 401 Ser 401  
Db 1225 TCA 1227

## RESULT 12

AAH14513  
ID AAH14513 standard; cDNA; 2524 BP.  
XX  
AC AAH14513;  
AC  
XX 26-JUN-2001 (first entry)  
DE Human cDNA sequence SEQ ID NO:12044.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
OS  
XX EPI074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
XX 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
Ishii S, Sugiyama T, Wakamatsu A, Nagai X, Otsuki T;  
WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID 12044; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH0166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2524 BP; 583 A; 747 C; 704 G; 490 T; 0 other;

Alignment Scores:  
Pred. No.: 1 67e-46 Length: 2524  
Score: 571.00 Matches: 145











GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 30, 2004, 03:41:40 ; Search time 4828 Seconds  
(without alignments)  
3397.838 Million cell updates/sec

Title: US-09-729-674-2  
Perfect score: 2130  
Sequence: 1 MEGORGAALLALLCVACA.....DDEGEVVEYVDDLELEETS 401

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ \_p2n.model -DEV=xlh  
-Q=/cgn2.1/USPTO.spool/US09729674/runat\_29012004\_095314\_12887/app\_query.fasta\_1.583  
-DB=GenEmbl -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=75 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09729674@cgn\_1\_3508 @runat\_29012004\_095314\_12887 -NCFU=6 -ICPU=3  
-NO MAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_em:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_on:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sv:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2128	99.9	2307	9	HSAJ6470	AJ006470 Homo sapi
2	2128	99.9	3871	6	BD140451	BD140451 Secreted
3	2125	99.8	1439	9	BC008745	BC008745 Homo sapi
4	2121	99.6	1961	6	BD156517	BD156517 Primer fo
5	2121	99.6	1961	9	AK001634	AK001634 Homo sapi
6	2101	98.6	2010	9	AK075401	AK075401 Homo sapi
7	2049.5	96.2	1922	9	AK091772	AK091772 Homo sapi
8	1986.5	93.3	1903	9	AK092990	AK092990 Homo sapi
9	1894	88.9	1885	10	MMAJ8469	AU008469 Mus muscu
10	1893	88.9	1863	10	BC049890	BC049890 Mus muscu
11	1827.5	85.8	1810	9	AK097850	AK097850 Homo sapi
12	1190.5	55.9	1665	5	GGCARPR	X97607 G.gallus mR
13	1155	54.2	818	6	BD148800	BD148800 Primer fo
14	1103.5	51.8	1903	10	BC031856	BC031856 Mus muscu
15	1090.5	51.2	2079	9	HSA250583	AJ750583 Homo sapi
16	1089.5	51.2	2037	9	BC001047	BC001047 Homo sapi
17	1089.5	51.2	2443	6	AX714040	AX714040 Sequence
18	1089.5	51.2	2443	9	AK056085	AK056085 Homo sapi
19	1088.5	51.1	1745	9	BC011701	BC011701 Homo sapi
20	1088.5	51.1	1745	9	BC011701	BC011701 Homo sapi
21	1088.5	51.1	2347	6	AX335490	AX335490 Sequence
22	1088.5	51.1	2347	9	HSU47821	X47821 Homo sapien
23	1027	48.2	1407	10	RNSC65MR	X65454 R.norvegicu
24	831	39.0	182509	9	AC112211	AC112211 Homo sapi
25	791	37.1	64706	2	AC136358	AC136358 Homo sapi
26	675.5	31.7	226615	2	AC122652	AC122652 Rattus no
27	627.5	29.5	176789	2	AC032013	AC032013 Mus muscu
28	627.5	29.5	217092	10	AL590968	AL590968 Mouse DNA
29	611	28.7	226822	2	AC112375	AC112375 Rattus no
30	611	28.7	241607	2	AC113836	AC113836 Rattus no
31	591	27.7	178345	9	AC091172	AC091172 Homo sapi
32	591	27.7	196735	2	AC012192	AC012192 Homo sapi
33	572.5	26.9	2993	6	BD158280	BD158280 Primer fo
34	572.5	26.9	2993	9	AK027648	AK027648 Homo sapi
35	571	26.8	2524	6	BD156505	BD156505 Primer fo
36	571	26.8	2524	9	AK027680	AK027680 Homo sapi
37	571	26.8	2563	6	AX136191	AX136191 Sequence
38	571	26.8	2563	9	BD123548	BD123548 Secretary
39	571	26.8	2563	9	AK075418	AK075418 Homo sapi
40	571	26.8	2583	6	AX704765	AX704765 Sequence
41	571	26.8	2585	6	BD156581	BD156581 Primer fo
42	571	26.8	2585	9	AK027697	AK027697 Homo sapi
43	567	26.6	2800	9	AF097432	AF097432 Homo sapi
44	565.5	26.5	2880	10	AF456412	AF456412 Mus muscu
45	545	25.6	2569	10	BC024047	BC024047 Mus muscu

ALIGNMENTS

RESULT 1









```

Db      1056 CACAGGACACGTGGGGCTCTCAGATGAGCACTTCCAGCCAGACCTGAAGCAGTTCAG 1115
Qy      361 PhePheAenValThrThLeuGlnLysGluLeuTyrAspPheAlaLysGluAenIleMet 380
Db      1116 TTCTTTAATGTGACCACACTCCAGAGGAGCTGTATGACTTTGCTAAGGAAAATATATG 1175
Qy      381 AspAspAspGluGlyValValGluTyrValAspAspLeuLeuGluGluThr 400
Db      1176 GATGATGATGAGGAGAAATTTGGATATGTGGATGACCTCTTGGAACTCGAGGAGACC 1235
Qy      401 Ser 401
Db      1236 AGC 1238

RESULT 4
BD156517
LOCUS      BD156517                      1961 bp      DNA      linear      PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION  BD156517
VERSION    BD156517.1  GI:27862275
KEYWORDS  JP 2002191363-A/11360.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1961)
AUTHORS   Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE     Primer for synthesizing full-length cDNA and use thereof
JOURNAL   Patent: JP 2002191363-A 11360 09-JUL-2002;
          HELIX RESEARCH INSTITUTE
COMMENT   OS Homo sapiens (human)
          PN JP 2002191363-A/11360
          PD 09-JUL-2002
          PF 28-JUL-2000 JP 2000280990
          PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
          PI SAITO,
          PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
          PI KEIICHI NAGAI, TETSUJI OTSUKI
          PC
          C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
          10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
          Primer for synthesizing full-length cDNA and use thereof PH Key
          Location/Qualifiers
          FT CDS      Location/Qualifiers
          1..1961
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
BASE COUNT  417 a 579 c 481 g 484 t
ORIGIN

Alignment Scores:
Pred. No.:      2.8e-176      Length:      1961
Score:          2121.00      Matches:      399
Percent Similarity: 99.50%      Conservative: 0
Best Local Similarity: 99.50%      Mismatches: 2
Query Match:    99.58%      Indels:      0
DB:            6            Gaps:        0

US-09-729-674-2 (1-401) x BD156517 (1-1961)
Qy      1 MetGluProGlyArgArgGlyValAlaAlaLeuLeuAlaLeuLeuCysValAlaCysAla 20
Db      40 ATGAGCGCGGGCGCGGGGGCGCGCGCTGCTAGCTGTGTCGGTGGCTCGCGG 99
Qy      21 LeuArgAlaGlyArgAlaGlnTyrGluArgTyrSerPheArgSerPheProArgAspGlu 40
Db      100 CTGGCGCGGGCGCGCGCAATACGAACGCTACAGCTTCCGAGCTTCCACGGAGCAG 159
Qy      41 LeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGlyGluHisTrp 60

```

```

Db      160 CTGATCCGCTCAGTCCGCTACCGGCACGCGCTGGCAGTACAGCGCGCAGCACTGG 219
Qy      61 AlaGluSer**GlyTyrLeuGluLysSerLeuAtcLeuHisArgLeuLeuArgAspSer 80
Db      220 GCCGAGAGCGTGGGCTTACCTGGAGATCAGCTGCGGCTGACCGCTTGTGTGGCGACG 279
Qy      81 GluAlaPheCysHisArgAsnCysserAlaAlaProGlnProGluProAlaAlaGlyLeu 100
Db      280 GAGGCCCTTCTGCCACCGCAACTGCAGCGCGCGCGCGCAGCGCGCGCGCGCGCGCTC 339
Qy      101 AlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArgArgAlaHisCysLeuLys 120
Db      340 GCCAGCTATCCGAGCTGCGCTCTTTCGGGGGCGCTCTGCGCGCGCGCTGCTCAAG 399
Qy      121 ArgCysLysGlnGlyLeuProAlaPheArgGlnSerGlnProSerArgGluValLeuAla 140
Db      400 CGCTGCAAGCAGGCGCTGCGAGCCCTCCGCCAGTCCCGAGCCCGCGCGCGGTGGCG 459
Qy      141 AspPheGlnArgArgGluProTyrLysPheLeuGlnPheAlaTyrPheLysAlaAsnAsn 160
Db      460 GACTTCCAGCGCGCGAGCCCTACAGGTTCCTGCAGTTCGCTTACTTCAAGCAATAAT 519
Qy      161 LeuProLysAlaAlaAlaAlaHisThrPheLeuLeuLysHisProAspAspGluMet 180
Db      520 CTCGCCAAGCCATCGCCGCTGCTCACCTTCTTACTGAAGCATCCTGATCGAAATG 579
Qy      181 MetLysArgAsnMetAlaTyrTyrLysSerLeuProGlyValaGluAspTyrLysAsp 200
Db      580 ATGAAGAGCAACATGCATATTATAAGAGCCCTGCTGGTGGCGAGGACTACATTAAAGAC 639
Qy      201 LeuGluThrLysSerTyrGluSerLeuPheLeuPheLeuAlaValArgAlaTyrAsnGly 220
Db      640 CTGGAACCAAGTCATATGAAGCCCTGTTCATCCGAGCAGTCCGGGCATACAAACGCTGAG 699
Qy      221 AsnTrpArgThrSerIleThrAspMetGluLeuAlaLeuProAspPhePheLysAlaPhe 240
Db      700 AACTGGAGAACATCCATCACACATGGAGCTGGGCTTCCGACTTCTTCAAAGCCCTTT 759
Qy      241 TyrGluCysLeuAlaAlaCysGluGlySerArgGluLysLysAspPheLysAspPheTyr 260
Db      760 TAGCAGTGTCTCGAGCCTCGAGGCTTCAGGGAGATCAAGGACTTCAAGGATTCTTAC 819
Qy      261 LeuSerIleAlaAspHisTyrValGluValLeuGluCysLysLysLysGlnCysGluLys 280
Db      820 CTTTCCATAGCAGATCATTTAGTAAGTCTTGGAAATGCCAAATACAGTGTGAAGAAC 879
Qy      281 LeuThrProValIleGlyTyrProValGluLysPheValAlaThrMetTyrHistyr 300
Db      880 CTCACCCAGTTATAGGAGGCTATCCGCTTGAGAAATTTGTGGCTACCATGTATCATTAC 939
Qy      301 LeuGlnPheAlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAlaProCysAlaValSer 320
Db      940 TTGCAGTTTGCTATTATAAGTTGAACCATCGCTGAAGATGCGAGCCCTGTGCGATCAGC 999
Qy      321 TyrLeuLeuPheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyrTyrGlnTyr 340
Db      1000 TATCTGCTTTTGATCAAGATGACAGGCTCATGCGAGCAGAACCTTGGTGTATACAGTAC 1059
Qy      341 HisArgAspThrTrpGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGln 360
Db      1060 CACAGGACACGTGGGGCTCTCAGATGAGCACTTCCAGCCCGAGCCCTGAAGCAGATT 1119
Qy      361 PhePheAsnValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGluAsnIleMet 380
Db      1120 TTCCTTAATGTGACCACTCCAGAGGAGCTGTATGACTTTGTGAAGAAAATATATATG 1179
Qy      381 AspAspAspGluGlyGluValValGluTyrValAspAspLeuLeuGluGluThr 400
Db      1180 GATGATGATGAGGAGAGAGTTGTGGAATATGTGATGACCTTCTTGGACTTGGAGAGACC 1239
Qy      401 Ser 401

```















204 LysSerTyrGluSerLeuPheIleArgAlaValArgAlaTyrAsnGlyGluAsnTyrArg 223  
 Db AAGTCGTACAGAGCGCTGTTGTCGCGGGCGCGCTCAACAGCGGAGACTGGAGA 716  
 224 ThrSerIleThrAspMetGluLeuAlaLeuProAspPhePheLeuAlaPheTyrGluCys 243  
 Db AGCTCCATTTCGAGATGAGCTCGGCTTCCCGACTTCCTCAAGGCGCTTACGAGTGC 776  
 244 LeuAlaAlaCysGluGlySerArgGluIleLysAspPheLysAspPheTyrLeuSerIle 263  
 Db CTGGCTGCTCGAGGGTTCGGGAGATCAAGGACTTCAGGACTTCTACTCTCCATA 836  
 264 AlaAspHisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsnLeuThrPro 283  
 Db CGAGATCACTATGTGGAAGTTCGTGAGTGTAGATTCGTTGTGAGGAGACCTCCACCCA 896  
 284 ValIleGlyGlyTyrProValGluLysPheValAlaThrMetTyrHisTyrLeuGlnPhe 303  
 Db GTTCATAGGAGGTATCCGCTGGGGAATTTGTGGCACCACTGACCACTATTTACAGTTT 956  
 304 AlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAlaProCysAlaValSerTyrLeuLeu 323  
 Db GCGTATTACAGTTCAATGATCTGAAGATCGAGCCCGCTGCGCTCAGCTACTCTGCTC 1016  
 324 PheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyrGlnTyrHisArgAsp 343  
 Db TTTGACCGAGATGACGGTCTATGACAGAACCTGGTGTACTATCATGATCCACCGGAC 1076  
 344 ThrTrpGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGlnPheAsn 363  
 Db AAGTGGGCGCTTCGATGAGCATTCCAGCCAGACCCGAGCAGTTCAGTTCTTTAAT 1136  
 364 ValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGluAsnIleMetAspAsp 383  
 Db GTGACGAGCTCCAGAGAACTGTACGACTTCGCTCAGGAACCTTAATGGATGACGAT 1196  
 384 GluGlyGluValValGluTyrValAspAspLeuLeuGluGluThrSer 401  
 Db GAGGAGAGGTTGTGAGTATGTGACGACTTGTGGAGACGGAAGTCTGGCC 1250  
 RESULT 11  
 AK097650  
 LOCUS  
 DEFINITION Homo sapiens cDNA FLJ40331 fis, clone TEST2031687, highly similar to Homo sapiens mRNA for cartilage-associated protein.  
 ACCESSION AK097650  
 VERSION AK097650.1 GI:21757488  
 KEYWORDS cDNA capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 Kanehori, K., Ishibashi, T., Chiba, Y., Fujimori, K., Hiraoka, S., Tanai, H., Watanabe, S., Ishida, S., Ono, Y., Hotta, T., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K., and Isogai, T.  
 NEO human cDNA sequencing project  
 JOURNAL Unpublished  
 REFERENCE 2 (Bases 1 to 1810)  
 AUTHORS Isogai, T. and Yamamoto, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Karusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 COMMENT NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology (RAB); cDNA library

construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
 Location/Qualifiers  
 1. 1810  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="TESTI2031687"  
 /cissue\_type="testis"  
 /clone\_lib="TESTI2"  
 /notes="cloning vector: pVE18SFJ3"

## FEATURES

## source

BASE COUNT 375 a 550 c 449 g 436 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 1348-150 Length: 1810  
 Score: 1827.50 Matches: 355  
 Percent Similarity: 88.78% Conservative: 1  
 Best Local Similarity: 88.53% Mismatches: 2  
 Query Match: 85.80% Indels: 44  
 DB: 9 Gaps: 1

US-09-729-674-2 (1-401) x AK097650 (1-1810)

QY 1 MetGluProGlyArgGlyAlaAlaLeuLeuAlaLeuLeuCysValaCysAla 20  
 Db ATGAGCCGGGGCGCGGGCGCGCTGCTAGCGCTCTGTGGCTGGCTGGCGG 84  
 QY 21 LeuArgAlaGlyArgAlaGlnTyrGluArgTyrSerPheArgSerPheProArgGlu 40  
 Db CTGCGCGCGCGGGCGCGCCCAATACAGCTACAGCTTCCGAGCTTCCCGGACGAG 144  
 QY 41 LeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGlyLysIleTyr 60  
 Db CTGATGCCCTCGAGTCGGCTTACCGGACGCGCTGGACAAAGTACAGCGGAGCACTGG 204  
 QY 61 AlaGluSer\*\*GlyTyrLeuGluLeuSerLeuArgLeuHisArgLeuLeuArgAspSer 80  
 Db GCGGAGAGCGTGGGCTACTCGAGATCAGCTCGCGCTGCGCGCTTCTCGCGGACAGC 264  
 QY 81 GluAlaPheCysHisArgAsnCysSerAlaAlaProGlnProGluProAlaAlaGlyLeu 100  
 Db GAGGCTTCTGCCACCGCAACTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 324  
 QY 101 AlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArgAlaHisCysLeuLys 120  
 Db GCCAGCTATCCGAGCTGCGCTCTTCGGGGGCGCTGCTGCGCGCGCGCTGCTCAAG 384  
 QY 121 ArgCysLysGlnGlyLeuProAlaPheArgGlnSerGlnProSerArgGluValLeuAla 140  
 Db CCGTGCAGAGCGGCGCTGCCAGCTTCCGCCAGTCCCGACCGCGCGCGCGCTGCTGGCG 444  
 QY 141 AspPheGlnArgArgGluProTyrTyrLysPheLeuGlnPheAlaTyrPheLysAlaAsn 160  
 Db GACTTCCAGCGCGCGGAGCCCTTACAAGTCTCTCGAGTTCGCTTACTTCAAGGCAAAAT 504  
 QY 161 LeuProLysAlaIleAlaAlaAlaHisThrPheLeuLeuLysHisProAspAspGluMet 180  
 Db CCCCCAAAGCCATCGCGCTGCTCACACCTTCTTACTGAAGCATCTCGATGAGGAAATG 564  
 QY 181 MetLysArgAsnMetAlaTyrTyrLysSerLeuProGlyAlaGluAspTyrIleLysAsp 200  
 Db ATGAAGAGAGCAATGCGCATATTATAAGAGCGCTGCTGGTGGCGGAGGACTACTT-AAAGAC 623  
 QY 201 LeuGluThrLysSerTyrGluSerLeuPheIleArgAlaValArgAlaTyrAsnGlyGlu 220  
 Db CTGGAACCAAGTCAATATGAAGCGCTGTTCATCCGAGCAGTGGCGGCATACCAACGGTGAG 683  
 QY 221 AsnTyrArgThrSerIleThrAspMetGluLeuAlaLeuProAspPhePheLysAlaPhe 240

```

Db 684 AACTGGAGACATCCATCACACATGAGCTGGCCCTCCGACTTCTTCAAGCCCTT 743
QY 241 TyrGluCysLeuAlaAAlaCysGluGlySerArgGluLeuLysAspPheLysAspPheTyr 260
Db 744 TACGAGTGTCTCGACGCTTCGAGGGTTCACAGGAGATCAAGGACTTCAAGGATTTCTAC 803
QY 261 LeuSerIleAlaAspHisTyrValGluValLeuGluCysLysIleGlnCysGluGluAsn 280
Db 804 CTTTCCATAGCA----- 815
QY 281 LeuThrProValIleGlyGlyTyrProValGluLysPheValAlaThrMetTyrHisTyr 300
Db 815 ----- 815
QY 301 LeuGlnPheAlaTyrTyrLysLeuAsnAspLeuLysAsnAlaAlaProCysAlaValSer 320
Db 816 -----GTGAACACCTGAGAAATGACGCCCTGTGCAGTCAGC 854
QY 321 TyrLeuLeuPheAspGlnAsnAspLysValMetGlnGlnAsnLeuValTyrTyrGlnTyr 340
Db 855 TATCTGCTCTTGAATCAGATGACAAAGTTCATGAGCAGAACTTGGTGTATTACAGTAC 914
QY 341 HisArgAspThrTyrGlyLeuSerAspGluHisPheGlnProArgProGluAlaValGln 360
Db 915 CACAGGACACGCTGGGGCTCTCAGATGAGCACTTCAGCCGACAGCTTGAAGCAGTTTAC 974
QY 361 PhePheAsnValThrThrLeuGlnLysGluLeuTyrAspPheAlaLysGluAsnIleMet 380
Db 975 TCTTTATGTGACCACTCAGAGGAGCTGTATGACTTTGTCTAAGGAAATATATATG 1034
QY 381 AspAspGluGlyGluValValGluTyrValAspAspLeuLeuGluGluThr 400
Db 1035 GATGATGATGAGGAGAAATTTGGAATATGTGATGACCTTTTGAACCTGGAGAGACC 1094
QY 401 Ser 401
Db 1095 AGC 1097

RESULT 12
GCGARPR 1665 bp mRNA linear VRT 26-JUN-1997
LOCUS G.gallus mRNA for cartilage associated protein.
DEFINITION X97607 GI:1296525
ACCESSION cartilage associated protein.
VERSION Gallus gallus (chicken)
KEYWORDS Gallus gallus
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1
AUTHORS Castagnola, P., Gennari, M., Morello, R., Tonachini, L., Marin, O.,
Gaggero, A. and Cancedda, R.
TITLE Cartilage associated protein (CASP) is a novel developmentally
regulated chick embryo protein
J Cell. Sci. 110 (Pt 12), 1351-1359 (1997)
J9760293
MEDLINE 97360293
PUBMED 9217321
AUTHORS Castagnola, P.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1996) P. Castagnola, Centro di Biotecnologie
Avanzate, Differenziamento Cellulare, L.go R. Benzi, 10 Genova,
16132, ITALY
FEATURES
Location/Qualifiers
1..1665
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/cell_type="chondrocyte"
/lab_stage="embryo"
/lab_host="E.coli"
43..858
CDS

```

```

/codon_start=1
/product="cartilage associated protein"
/protein_id="CAA66206.1"
/db_xref="GI:1296526"
/db_xref="SWISS-PROT:Q90830"
/translation="NMRLAALLATAGAQYERYFRFPRLDMLPESAYRYGLDQY
STENPESVYLEVSRMLRRLURDTEAFCHNCSSAGPLTAPPDAGELEAURLLAGV
LRRACLRCKQGLFAQAFQRELEEFQRREPKYLFQFAFYKXANLPKAAAH
FLIKDDMMQRMVAYKSIDABEHKDLREKVENLFVRAVAYNGNWRISID
MELALPDPFFKTDGCIACGEGSRKDKDFYLSIADHYIEVLACKVQPDIT"
BASE COUNT 470 a 389 c 388 g 418 t
ORIGIN
Alignment Scores: 8.16e-95 Length: 1665
Pred. No.: 1190.50 Matches: 238
Score: 68.15% Conservative: 38
Best Local Similarity: 58.77% Mismatches: 36
Query Match: 55.89% Indels: 94
DB: 5 Gaps: 6
US-09-729-674-2 (1-401) x GCGARPR (1-1665)
QY 3 ProGlyArgArgGlyAla-----AlaAlaLeuLeuAlaLeuLeuVal 17
Db 4 CCACGGCGACGTGGTCCGGCGGCGGCGGAGGATGTGGCGGACGTGTGGCG 63
QY 18 AlaCysAlaLeuArgAlaGlyArgAlaGlnTyrGluArgTyrSerPheArgSerPhePro 37
Db 64 GCGCTTCTGGCGACGGCGGCG---GCGCAGTACGAGCGGTACAGCTTCGCGAGCTCCCG 120
QY 38 ArgAspGluLeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGly 57
Db 121 CGGACGAGCTGATGCCCTGGAATCCGCTACCGCTACGGCTTGACGACGACGACG 180
QY 58 GluHisTrpAlaGluSer***GlyTyrLeuGluIleSerLeuArgLeuHisArgLeuLeu 77
Db 181 GAGAACTGCGCGAGAGCGTCAGCTACCTGAGGAGTGGAGTGGCGGTGTACCGCTGCTG 240
QY 78 ArgAspSerGluAlaPheCysHisArgAsnCysSerAlaAla-----ProGlnPro 94
Db 241 CGCGACACGAGGCGCTTTCGCCACCACTCGCGCTCGCGCGCGCGCGCTCACCGCCCG 300
QY 95 GluProAlaAlaGlyLeuAlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArg 114
Db 301 CGCGCGCGCGATGGG-----CAGCTGCGCGAGCTGCGCTGCTGGCGGAGTGTGGCG 354
QY 115 ArgAlaHisCysLeuLysArgCysLysGlnGlyLeuProAlaPheArgGlnSerGlnPro 134
Db 355 CGAGCGCAATGCTGCGCGCTGCAAGCAGGCGGTTCGCGCGCTTCGCGCGCGCGCGCC 414
QY 135 SerArgGluValLeuAlaAspPheGlnArgGluProTyrLysPheLeuGlnPheAla 154
Db 415 GCGCGCGAGCTGCTCGAGAGTTCGAGCGCGCGAGAGCGCTTCAAGTACCTCAGTTCCGC 474
QY 155 TyrPheLysAlaAsnLeuProLysAlaIleAlaAlaHisThrPheLeuLeuLys 174
Db 475 TACTTCAAGGCTATAAAGCTTCCAAAGCTATTGAGCAGCTCACACATTTCTTCTGAA 534
QY 175 HisProAspAspGluMetLysArgAsnMetAlaTyrTyrLysSerLeuProGlyAla 194
Db 535 CATCCAGATGATGAATGATCAAGAAATATGCGCTTACTATAAGAGATACCTGATGCT 594
QY 195 GluAspTyrIleLysAspLeuThrLysSerTyrGluSerLeuPheIleArgAlaVal 214
Db 595 GAGGAGCATTAAAGACTTGGAGCAAGAGCTTACGAGATCTTCTTGTGAGGGCTGTG 654
QY 215 ArgAlaTyrAsnGlyGluAsnTrpArgThrSerIleThrAspMetGluLeuAlaLeuPro 234
Db 655 AGAGCATACAATGGTGTGACAACTGGAGAACATCCATCTCAGATATGGAATCGCTTCT 714
QY 235 AspPhePheLysAlaPheTyrGluCysLeuAlaAlaCysGluGlySerArgGluLeuLys 254

```

```

Db 715 GATTTCCTCAAACTACAGACTGTATAGCAGCGCTGTGAGGCTCCGAGAGATTAAA 774
QY 255 AspPheLysAspPheTyrLeuSerIleAlaAspHisTyrValGluValLeuGluCysLys 274
Db 775 GATTTCAGAGCTTCTATCTCTATTCAGAGATCATATTATGAAGTCTTGCATGCATA 834
QY 275 IleGlnCysGluGluAsnLeuThrProValIleGlyGlyTyrProValGluLysPheVal 294
Db 835 GTGCAG----- 840
QY 295 AlaThrMetTyrHisTyrLeuGlnPheAlaTyrTyrLysLeuAsnAspLeuLysAsnAla 314
Db 840 ----- 840
QY 315 AlaProCysAlaValSerTyrLeuLeuPheAspGlnAsnAspLysValMetGlnGlnAsn 334
Db 840 ----- 840
QY 335 LeuValTyrTyrGlnTyrHisArgAspThrTrpGlyLeuSerAspGluHisPheGlnPro 354
Db 841 -----ITCGA-TATCAC----- 851
QY 355 ArgProGluAlaValGlnPhePheAsnValThrLeuGlnLysGluLeuTyrAspPhe 374
Db 852 -----AACATAACCCACACTCCAGTTGGAAATGTATGAATTC 887
QY 375 AlaLysGluAsnIleMetAspAspAspGluGlyGluValValGluTyrValAspAspLeu 394
Db 888 GCAAGGAAACATCTTATGATGATGATGAGGTGAAGTTGTGAATTCCTTGATGAGTTG 947
QY 395 LeuGluLeuGluGlu 399
Db 948 TTAGAAATGGAGAA 962

RESULT 13
BD148800
LOCUS
DEFINITION
Primer for synthesizing full-length cDNA and use thereof.
ACCESSION
BD148800
VERSION
BD148800.1 GI:27854558
KEYWORDS
JP 2002191363-A/3643.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 818)
Ota.T., Isogai.T., Nishikawa.T., Hayashi.K., Saito.K., Yamamoto.J.,
Ishii.S., Sugiyama.T., Wakamatsu.A., Nagai.K. and Otsuki.T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 3643 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/3643
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
I SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source 1. 818
FT Location/Qualifiers
/organism='Homo sapiens (human)'.
1. 818
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
BASE COUNT 159 a 271 c 236 g 148 t 4 others

```

## ORIGIN

Alignment Scores: 4,11e-92 Length: 818  
 Pred. No.: 1155.00 Matches: 238  
 Score: 91.57% Conservative: 1  
 Best Local Similarity: 91.19% Mismatches: 15  
 Query Match: 54.23% Indels: 8  
 DB: 6 Gaps: 1

US-09-729-674-2 (1-401) x BD148800 (1-818)

```

QY 1 MetGluProGlyArgArgGlyAlaAlaAlaLeuLeuAlaLeuLeuCysValAlaCysAla 20
Db 40 ATGAGCGCGGGCGCGGGGGCGGGGGCGCTGCTAGCGCTGCTGCTGCTGCTGCTGCTGCG 99
QY 21 LeuArgAlaGlyArgAlaGlnTyrGluArgTyrSerPheArgSerPheProArgaspGlu 40
Db 100 CTGCGCGCGGGCGCGGGCGGGCGGGCGCTAGCTTCCGAGCTTCCACGGGACGAG 159
QY 41 LeuMetProLeuGluSerAlaTyrArgHisAlaLeuAspLysTyrSerGlyGluHisTrp 60
Db 160 CTGATCGCGCTCGAGTCGGCTACCGCGCACGGCTGGCAAGTACAGCGCGGACACTGG 219
QY 61 AlaGluSer***glyTyrLeuGluIleSerLeuArgLeuHisArgLeuLeuArgaspSer 80
Db 220 GCCGAGAGCGGTGGCTGACCTGGAGATCAGCTCGCGCTGACCGCTTGTGCGCGACAGC 279
QY 81 GluAlaPheCysHisArgAsnCysSerAlaAlaProGlnProGluProAlaAlaGlyLeu 100
Db 280 GAGGCTTCTGCCACCGCAACTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 339
QY 101 AlaSerTyrProGluLeuArgLeuPheGlyGlyLeuLeuArgAlaHisCysLeuLys 120
Db 340 GCCAGCTATCCGAGCTGCGCTCTTCGGGGGCGCTTCTCGCGCGCGCGCGCGCTCAAG 399
QY 121 ArgCysLysGlnGlyLeuProAlaPheArgGlnSerGlnProSerArgGluValLeuAla 140
Db 400 CGCTGCAAGCAGCGGCTTCCAGCTTCCGCGAGTCCAGCCCGCGCGCGCGGCTGGGG 459
QY 141 AspPheGlnArgArgGluProTyrLysPheLeuGlnPheAlaTyrPheLysAlaAsn 160
Db 460 GAUCCAGCGCGCGCGCGCGCGCTTCCAGTTCCTGCGAGTTCCTTACTTCAAGCAATAAT 519
QY 161 LeuProLysAlaIleAlaAlaHisThrPheLeuLeuLysHisProAspAspGluMet 180
Db 520 CTCGCCAAGCCATCGCCCTGCTCACCTTTCTACTGAAGCATCCTGATCAGCAATG 579
QY 181 MetLysArgAsnMetAlaTyrTyrLysSerLeuProGlyValaGluAspTyrIleLysAsp 200
Db 580 ATGAGAGAGAAACATGGCATATTATAAGAGCTGCT-GGTGCGCGAGGACTTACATTAAAG 638
QY 201 LeuGluThrLys-SerTyrGlu-SerLeuPheIleArgAlaValaArgAlaTyrAsnGly 220
Db 639 CTGGAAACCAAGTTCATATGAAGAGCTTTTATCCGANCAGTGGCGGATACAAACGGTG 698
QY 220 LuAsnTrpArgThrSerIleThrAspMetGluLeuAla--LeuProAspPhePhe-LysA 239
Db 699 AAAATGGAGAACATTCATCACAGCATGGAAGTGGCCCTTTCCGAAATTTCTTCAAAG 758
QY 239 lapherTyrGlu-----CysLeuAlaAlaCysGluGlySerArgGluLys 254
Db 759 NCTTTTACGAATGGTCTTCANCCCTGGCGAAGGGTTNCAAGGGAAGAAATCAAG 811

```

## RESULT 14

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

BC031856 1903 bp mRNA linear ROD 22-APR-2003  
 Mus musculus SC65 synaptonemal complex protein, mRNA (cDNA clone  
 MGC:25554 IMAGE:3968941), complete cds.  
 BC031856  
 BC031856.1 GI:21594429  
 MGC.  
 Mus musculus (house mouse)









GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 07:10:40 ; Search time 4381.42 Seconds  
(without alignments)  
11232.506 Million cell updates/sec

Title: US-09-729-674-1\_COPY\_63\_1265  
Perfect score: 1203  
Sequence: 1 atgagccg99gcccgggg.....tggaaactggaggagaccagc 1203

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_by.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_nam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1202.2	99.9	2307	9	HSJ6470
2	1202.2	99.9	3871	6	BD140451
3	1197.4	99.5	1439	9	BC008745
4	1197.4	99.5	1361	6	BD156517
5	1197.4	99.5	1961	9	AK001634
6	1187	98.7	2010	9	AK075401
7	1111.4	92.4	1922	9	AK091772
8	1076.6	89.5	1903	9	AK092990
9	923.8	76.8	1663	10	BC049890
10	923.8	76.8	1685	10	MMAJ6469
11	919	76.4	1810	9	AK097650
12	656.6	54.6	818	6	BD148800
13	496.2	41.2	1665	5	GGCARPR1
14	475.6	39.5	182509	9	AC112211
15	467	38.8	1889	9	BC001047
16	467	38.8	2037	9	BC007942
17	467	38.8	2443	6	AX714040
18	467	38.8	2443	9	AK056085
19	466.6	38.8	1745	9	BC011701
20	465.8	38.7	2347	6	AX335490
21	465.8	38.7	2347	9	HSU47621
22	465.4	38.7	2079	9	HSJ250583
23	448.2	37.3	64706	2	AC136358
24	428.2	35.6	1903	10	BC031856
25	403	33.5	1407	10	RNSC65MR
26	347.4	28.9	226615	2	AC122652
27	205.2	17.1	178345	9	AC091172
28	205.2	17.1	196735	2	AC012132
29	179.2	14.9	176789	2	AC032013
30	179.2	14.9	217032	10	AL590968
31	174.4	14.5	226822	2	AC112575
32	174.4	14.5	241607	6	BD136336
33	156.4	13.0	2524	6	BD156505
34	156.4	13.0	2524	9	AK027680
35	156.4	13.0	2563	6	AX136191
36	156.4	13.0	2563	6	BD123548
37	156.4	13.0	2563	9	AK075418
38	156.4	13.0	2583	6	AX704765
39	156.4	13.0	2585	6	BD156361
40	156.4	13.0	2585	9	AK027697
41	156.4	13.0	2993	6	BD158280
42	156.4	13.0	2993	9	AK027648
43	156	13.0	2249	10	MMU430350
44	154.8	12.9	2600	9	AF097432
45	154.8	12.9	2829	9	AF097431

ALIGNMENTS

RESULT 1	HSJ6470	2307 bp	mRNA	linear	PRI 23-MAR-2000
HSJ6470	LOCUS	2307 bp	mRNA	linear	PRI 23-MAR-2000
DEFINITION	Homo sapiens mRNA for cartilage-associated protein (CASP).				
ACCESSION	AJ006470				
VERSION	AJ006470.1	GI:3687321			
KEYWORDS	cartilage-associated protein; CASP.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1				
AUTHORS	Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.				
TITLE	Tonachini, L., Morello, R., Monticone, M., Skaug, J., Scherer, S.W., Cancedda, R. and Castagnola, P.				
	CDNA cloning, characterization and chromosome mapping of the gene				

encoding human cartilage associated protein (CRTAP)  
 Cytogenet. Cell Genet. 87 (3-4), 191-194 (1999)  
 20169181  
 PUBMED  
 10702664  
 2 (bases 1 to 2307)  
 Castagnola, P.  
 Direct Submission  
 Submitted (12-JUN-1998) Castagnola P., Biotechnologie in Oncologia,  
 Ist. Naz.le per la Ricerca sul Cancro, L.G. R. Benzi 10 Genoa,  
 16132, ITALY

## FEATURES

source

Location/Qualifiers

1..2307  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="Contig of pCPH2 and I.M.A.G.E. cloneID 1090695"  
 /tissue type="fetal brain"  
 /dev\_stage="19-23 weeks-old"  
 1..2307  
 /gene="casp"  
 12..1217  
 /gene="casp"  
 /codon\_start=1  
 /product="cartilage-associated protein (CASP)"  
 /protein\_id="CAA07054.1"  
 /db\_xref="GI:3687322"  
 /db\_xref="SPTREMBL:O75718"  
 /translation="MEPGRGAAALLALLVACALRAGRAQVRYSPRSPDELMP  
 ESAYHADKYGEHWASVGYLEISLRHLRLDSEAFCHRNCSAAPQPPAAGLAS  
 YPELRLFGELRLAKRCKQGLPFRSQSPREVADFORRPYKFLQFPAYKANN  
 LPKATAAHTFLKHPDDMMKRNMAVYKSLPGADYIKDLETYSYSLFTRAVYAN  
 GENWTSITDNEALPDRFKAFYECLACEGSRKIDKDYLSIADHYEVLCKIQ  
 CENITPVGIPVPEKVFATMHIYQFAYIKLNDKNAAPCAVSIFLFDQNDKVMQON  
 LVYQIHRDTWGLSDEHFQPRPEAVQFFNVITLQKELYDFAKENIMDDDEGEVVEYVD  
 DLLELEETS"  
 12..89  
 /gene="casp"  
 90..1214  
 /gene="casp"

sig\_peptide

mat\_peptide

BASE COUNT 545 a 633 c 548 g 578 t  
 ORIGIN

Query Match 99.9%; Score 1202.2; DB 9; Length 2307;  
 Best Local Similarity 99.8%; Pred. No. 2.6e-182;  
 Matches 1201; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGAGCCGGCGCGCGGGCGCGCGCTGCTAGCGCTGCTGTGCGTGCCTGCGCG 60  
 DB 12 ATGGAGCCGGCGCGCGGGCGCGCGCTGCTAGCGCTGCTGTGCGTGCCTGCGCG 71

QY 61 CTGGCGCGCGCGCGCGCGCAATACGACGCTACGCTCCGAGCTTCCACGGGACGAG 120  
 DB 72 CTGGCGCGCGCGCGCGCGCAATACGACGCTACGCTTCCGAGCTTCCACGGGACGAG 131

QY 121 CTGATGCGCGCTCGAGTTCGCGCTACCGGACGCGCTGGAAGTACAGCGCGGAGCACTGG 180  
 DB 132 CTGATGCGCGCTCGAGTTCGCGCTACCGGACGCGCTGGAAGTACAGCGCGGAGCACTGG 191

QY 181 GCGGAGAGCTGCGTACTGAGATCAGCTGCGGCTGCACCGCTTCTGCGGAGCAGC 240  
 DB 192 GCGGAGAGCTGCGTACTGAGATCAGCTGCGGCTGCACCGCTTCTGCGGAGCAGC 251

QY 241 GAGGCGCTTCTGCGACGCGCACTGCGCGCGCGCGAGCGCGCGCGCGCGCTC 300  
 DB 252 GAGGCGCTTCTGCGACGCGCACTGCGCGCGCGCGAGCGCGCGCGCGCGCTC 311

QY 301 GCGAGTATCCGAGCTGCGCTTTCGCGGCGCTGCTGCGCGCGCGCACTGCGCTCAAG 360  
 DB 312 GCGAGTATCCGAGCTGCGCTTTCGCGGCGCTGCTGCGCGCGCGCACTGCGCTCAAG 371

QY 361 CGCTGCAACGAGGCTGCGCGCTTTCGCGGCGCTGCGCGCGCGCGCGGAGTGTGCGG 420

Db 372 CGCTGCAAGCAGGCGCTGCCAGCCTTCCGCGAGTCCCGAGCCCGCGAGGTGCTGGCG 431

QY 421 GACTTCCAGCGCGCGAGCCCTACAGTTCCTGCGAGTTCCTGCTTACTTCAAGGCAATAAT 480

Db 432 GACTTCCAGCGCGCGAGCCCTACAGTTCCTGCGAGTTCCTGCTTACTTCAAGGCAATAAT 491

QY 481 CTCCCAAGCCATCCCGCTGCTCACCTTCTACTGAAGCATCCTGATGACGAAATG 540

Db 492 CTCCCAAGCCATCCCGCTGCTCACCTTCTACTGAAGCATCCTGATGACGAAATG 551

QY 541 ATGAAGAGGACATGGCATATTATATAGAGCTGCTGCTGCGGAGCACTACATTAAAGAC 600

Db 552 ATGAAGAGGACATGGCATATTATAGAGCTGCTGCTGCGGAGCACTACATTAAAGAC 611

QY 601 CTGGAACCAAGTCATATGAAGCCCTGTTTCATCCGAGCAGTGGCGGCATACACGGTGAG 660

Db 612 CTGGAACCAAGTCATATGAAGCCCTGTTTCATCCGAGCAGTGGCGGCATACACGGTGAG 671

QY 661 AACTGGAGACATCCCATCAGACATGGAGCTGGGCCCTTCCGACTTCTTCAAGCCCTT 720

Db 672 AACTGGAGACATCCCATCAGACATGGAGCTGGGCCCTTCCGACTTCTTCAAGCCCTT 731

QY 721 TAGCAGTGTCTCGCAGCCTCGAGGGTTCAGGAGATCAAGGACATTCAGGATTTCTAC 780

Db 732 TAGCAGTGTCTCGCAGCCTCGAGGGTTCAGGAGATCAAGGACATTCAGGATTTCTAC 791

QY 781 CTTTCCATAGCAGATCATTTATGTAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAAC 840

Db 792 CTTTCCATAGCAGATCATTTATGTAAGTTCTGGAATGCAAAATACAGTGTGAAGAGAAC 851

QY 841 CTCACCCCGTTATAGGAGCTATCCGTTGAGAAATTTGTGCTACCATGTATCATTAC 900

Db 852 CTCACCCCGTTATAGGAGCTATCCGTTGAGAAATTTGTGCTACCATGTATCATTAC 911

QY 901 TTGCAAGTTGCCCTATTATAAGTTGAAACGACCTGGAAGTGCAGCCCTGTGCAGTCAGC 960

Db 912 TTGCAAGTTGCCCTATTATAAGTTGAAACGACCTGGAAGTGCAGCCCTGTGCAGTCAGC 971

QY 961 TATCTGCTCTTGCATCAGATGACAGGTCATGACAGAACTGCTGCTGTTTACCACTAC 1020

Db 972 TATCTGCTCTTGCATCAGATGACAGGTCATGACAGAACTGCTGCTGTTTACCACTAC 1031

QY 1021 CACAGGACACACTTGGGGCTCTCGGATGAGCACTTCCAGCCCGAGACCTGGAAGCAGTTCAG 1080

Db 1032 CACAGGACACACTTGGGGCTCTCGGATGAGCACTTCCAGCCCGAGACCTGGAAGCAGTTCAG 1091

QY 1081 TTCTTTAATGTGACACACACTCCAGAGAGGCTGTATGACTTTTCTTAAGGAAATATAATG 1140

Db 1092 TTCTTTAATGTGACACACACTCCAGAGAGGCTGTATGACTTTTCTTAAGGAAATATAATG 1151

QY 1141 GATGATGATGAGGAGAGTGTGGAAATATGAGTACCTCTTGGAACTGGAGGAGACC 1200

Db 1152 GATGATGATGAGGAGAGTGTGGAAATATGAGTACCTCTTGGAACTGGAGGAGACC 1211

QY 1201 AGC 1203

Db 1212 AGC 1214

## RESULT 2

BD140451  
 LOCUS  
 DEFINITION  
 Secreted proteins and polynucleotides encoding them.  
 ACCESSION  
 BD140451  
 VERSION  
 BD140451.1 GI:23235396  
 KEYWORDS  
 JP 2002506611-A/1.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 3871)  
 Jacobs, K., McCoy, J.M., Lavallie, E.R., Racie, L.A.C., Evans, C.,  
 Merberg, D., Treacy, M., Agostino, M.J., Il, R.J.S., Wong, G.G.,

Clark, H.F. and Fechtel, K.  
 Secreted proteins and polynucleotides encoding them  
 Patent: JP 200206611-A 1 05-MAR-2002;  
 GENETICS INSTITUTE INC  
 COMMENT OS Homo sapiens (human)  
 PN JP 200206611-A/1  
 PD 05-MAR-2002  
 PF 24-NOV-1998 JP 2000522118  
 PR 26-NOV-1997 US 60/066804, 23-NOV-1998 US 09/197886 PI  
 KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A COLLINS PI  
 RACIE,  
 PI CHERYL EVANS, DAVID MERBERG, MAURICE TREACY, MICHAEL J AGOSTINO,  
 PI ROBERT J STEININGER II, GORDON G WONG, HILARY F CLARK, KIM PI  
 FECHTEL,  
 PC C12N15/09, C07K14/00, C12N1/21, C12N5/10, C12P19/34, C12P21/02, PC  
 C12Q1/68//  
 PC A61P29/00, A61P35/00, A61P37/04, A61P37/06, C12N15/00, C12N5/00 CC  
 Secreted proteins and polynucleotides encoding them FH Key  
 Location/Qualifiers  
 FT source 1..3871  
 FT /organism="Homo sapiens (human)".  
 FEATURES  
 source Location/Qualifiers  
 1..3871  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 965 a 913 c 1033 g 958 t 2 others  
 ORIGIN  
 Query Match 99.9%; Score 1202.2; DB 6; Length 3871;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-182;  
 Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGAGCGCGGCGCGCGCGCGCGCGCTGTAGCGCTGTGCGTGGCTGGCG 60  
 DB 63 ATGAGCGCGGCGCGCGCGCGCGCGCTGTAGCGCTGTGCGTGGCTGGCG 122  
 QY 61 CTGCGCGCGGCGCGCGCGCGCGCGCTGTAGCGCTGTGCGTGGCTGGCG 120  
 DB 123 CTGCGCGCGGCGCGCGCGCGCGCGCTGTAGCGCTGTGCGTGGCTGGCG 182  
 QY 121 CTGATCGCGCTGAGTCGGCTACCGGCACGCGCTGAGCAAGTACAGCGGCGGCGCTGG 180  
 DB 183 CTGATGCGCGCTGAGTCGGCTACCGGCACGCGCTGAGCAAGTACAGCGGCGGCGCTGG 242  
 QY 181 GCGGAGAGCTGCGCTGAGTACGCTGCGGCTGCGGCTGCGGCTGCGGCGAGC 240  
 DB 243 GCGGAGAGCTGCGCTGAGTACGCTGCGGCTGCGGCTGCGGCTGCGGCGAGC 302  
 QY 241 GAGGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 300  
 DB 303 GAGGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 362  
 QY 301 GCCAGCTATCGGAGCTCGCTCTTGGGGGCTGCTGCGCGCGCGCGCGCTCAAG 360  
 DB 363 GCCAGCTATCGGAGCTCGCTCTTGGGGGCTGCTGCGCGCGCGCGCGCTCAAG 422  
 QY 361 CGCTGCAAGCGGCGCTCGCGCGCTTCCGCGAGTCCGCGCGCGCGCGCGGCTGGCG 420  
 DB 423 CGCTGCAAGCGGCGCTCGCGCGCTTCCGCGAGTCCGCGCGCGCGCGGCTGGCG 482  
 QY 421 GACTTCCAGCGCGCGGAGCGCTTACAGTTCTGCGAGTTCGTTACTTCAAGCGCAATAT 480  
 DB 483 GACTTCCAGCGCGGAGCGCTTACAGTTCTGCGAGTTCGTTACTTCAAGCGCAATAT 542  
 QY 481 CTCGCCAAGCGCTCGCGCGCTGCTCACACCTTTCTACTGAGAGCTCTGTATGAGCAATG 540  
 DB 543 CTCGCCAAGCGCTCGCGCGCTGCTCACACCTTTCTACTGAGCAATCTGTATGAGCAATG 602  
 QY 541 ATGAGAGGAGCAATGCGCATATTAAGAGCGCTGCTGTCGCGGAGGACTCATTTAAGAC 600  
 DB 603 ATGAGAGGAGCAATGCGCATATTAAGAGCGCTGCTGTCGCGGAGGACTCATTTAAGAC 662

QY 601 CTGGAACCAAGTCATATGAAAGCCTGTTTCATCCGAGCAGTCGGGCATACAAACGGTGAG 660  
 DB 663 CTGGAACCAAGTCATATGAAAGCCTGTTTCATCCGAGCAGTCGGGCATACAAACGGTGAG 722  
 QY 661 AACTGGAGAACATCCATCACAGACATGGAGCTGGCCCTCCGAGCTTCTTCAAGGCTTT 720  
 DB 723 AACTGGAGAACATCCATCACAGACATGGAGCTGGCCCTCCGAGCTTCTTCAAGGCTTT 782  
 QY 721 TAGCAGTGTCTCGCAGCCTCGAGGGTTCAGGGGAGATCAAGGACTTCAAGGATTTCTAC 780  
 DB 783 TAGCAGTGTCTCGCAGCCTCGAGGGTTCAGGGGAGATCAAGGACTTCAAGGATTTCTAC 842  
 QY 781 CTTTCATGACAGATCATTTATGATAGAGTTCGGAATGCAAAATACAGTGTGAAGAGAC 840  
 DB 843 CTTTCATGACAGATCATTTATGATAGAGTTCGGAATGCAAAATACAGTGTGAAGAGAC 902  
 QY 841 CTCACCCCGAGTTATAGGAGGCTATCCGGTTGAGAAATTTGGCTACCATGATGATCATTAC 900  
 DB 903 CTCACCCCGAGTTATAGGAGGCTATCCGGTTGAGAAATTTGGCTACCATGATGATCATTAC 962  
 QY 901 TTGCGATTTGCCATTTATTAAGTTGAACGACCTGGAAGATGACGCCCTGTGCGAGTCAGC 960  
 DB 963 TTGCGATTTGCCATTTATTAAGTTGAACGACCTGGAAGATGACGCCCTGTGCGAGTCAGC 1022  
 QY 961 TATCTGCTCTTTGATCAGAAATGACAAAGTCTATGAGTGAAGTGAAGTGAAGTGAAGTGAAG 1020  
 DB 1023 TATCTGCTCTTTGATCAGAAATGACAAAGTCTATGAGTGAAGTGAAGTGAAGTGAAGTGAAG 1082  
 QY 1021 CACAGGACACTTGGGGCTCTCGGATGAGCACTTCCAGCCGAGACCTTCCAGGAGGAGC 1080  
 DB 1083 CACAGGACACTTGGGGCTCTCGGATGAGCACTTCCAGCCGAGACCTTCCAGGAGGAGC 1142  
 QY 1081 TTCTTTAATGTCACACACTCCAGAGGAGCTGATGACTTTTCTTAAGGAAATATAATG 1140  
 DB 1143 TTCTTTAATGTCACACACTCCAGAGGAGCTGATGACTTTTCTTAAGGAAATATAATG 1202  
 QY 1141 GATGATGATGAGGAGAGAGTTCGGAATATGATGAGTGAAGTTCGGAATCTGGAAGTGAAGAGAC 1200  
 DB 1203 GATGATGATGAGGAGAGAGTTCGGAATATGATGAGTGAAGTTCGGAAGTGAAGAGAC 1262  
 QY 1201 AGC 1203  
 DB 1263 AGC 1265  
 RESULT 3  
 BC008745  
 LOCUS  
 DEFINITION Homo sapiens, cartilage associated protein, clone MGC:1926  
 IMAGE:3347384, mRNA, complete cds.  
 ACCESSION BC008745  
 VERSION BC008745.1 GI:14250580  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1439)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: DCTD/DTP  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;





```
FEATURES
  source
    Location/Qualifiers
      1..1961
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="NT2RP4000243"
        /cell_line="NT2"
        /cell_type="teratocarcinoma"
        /clone_lib="NT2RP4"
        /notes="cloning vector: pME18SFL3-mRNA from NT2 neuronal
        precursor cells after 2-weeks retinoic acid (RA)
        induction."
      417 a 579 c 481 g 484 t

BASE COUNT 417 a 579 c 481 g 484 t
ORIGIN
  Query Match 99.5%; Score 1197.4; DB 9; Length 1961;
  Best Local Similarity 99.6%; Pred. No. 1.5e-181;
  Matches 1198; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGAGCCGGGCGCGGGGCGGGGCGGGGCGGCTGCTAGCCCTGCTGTGGTGGCTGGCGG 60
Db 40 ATGGAGCCGGGCGCGGGGCGGGGCGGGGCGGCTGCTAGCCCTGCTGTGGTGGCTGGCGG 99
QY 61 CTGCGGCGCGGCGCGGCCCAATACGAACGCTACAGCTTCGGCAGCTTCCACGGCAGCAG 120
Db 100 CTGCGGCGCGGCGCGGCCCAATACGAACGCTACAGCTTCCGAGCTTCCACGGCAGCAG 159
QY 121 CTGATCCGCTCGAGTCGGCTTACCGGACGGCTGGAAGTACAGCGGCGAGCACTGG 180
Db 160 CTGATCCGCTCGAGTCGGCTTACCGGACGGCTGGAAGTACAGCGGCGAGCACTGG 219
QY 181 GCGGAGGCTGGCTTACCTGAGATCAGCTTGGGCTGCGCTGCTGCTGCGGACAGC 240
Db 220 GCGGAGGCTGGCTTACCTGAGATCAGCTTGGGCTGCGCTGCTGCTGCGGACAGC 279
QY 241 GAGGCTTTCGACCGCAACTGCGAGCGCGCGCGCGCGAGCCCGCGCGCGCTC 300
Db 280 GAGGCTTTCGACCGCAACTGCGAGCGCGCGCGCGCGAGCCCGCGCGCGCTC 339
QY 301 GCGAGTATCCGAGTGGCTTCTCGGGGCGCTGCTGCGGCGCGCTGCTGCGGCGCTCAAG 360
Db 340 GCGAGTATCCGAGTGGCTTCTCGGGGCGCTGCTGCGGCGCGCTGCTGCGGCGCTCAAG 399
QY 361 CGCTGCAAGCAGGCGCTGCGGCTTCCGCGAGTCCCGAGCCCGCGGAGTGTGGCG 420
Db 400 CGCTGCAAGCAGGCGCTGCGGCTTCCGCGAGTCCCGAGCCCGCGGAGTGTGGCG 459
QY 421 GACTTCCAGCGCGGAGCCCTACAGTTCCTGCGAGTGGCTTCTTCAAGGCAATAT 480
Db 460 GACTTCCAGCGCGGAGCCCTACAGTTCCTGCGAGTGGCTTCTTCAAGGCAATAT 519
QY 481 CTCCCCAAGCCATCGCCGCTGCTCACACCTTCTACTGAAGCATCCTGTGACGAAATG 540
Db 520 CTCCCCAAGCCATCGCCGCTGCTCACACCTTCTACTGAAGCATCCTGTGACGAAATG 579
QY 541 ATGAAGAGAAATGCGATATTATTAAGAGCTGCTGCTGCGGAGGATCAATTAAAGAC 600
Db 580 ATGAAGAGAAATGCGATATTATTAAGAGCTGCTGCTGCGGAGGATCAATTAAAGAC 639
QY 601 CTGGAACCAAGTCATATGAAGCCGTGTCATCCGAGCAGTGGCGGATACAGCGGTGAG 660
Db 640 CTGGAACCAAGTCATATGAAGCCGTGTCATCCGAGCAGTGGCGGATACAGCGGTGAG 699
QY 661 AACTGAGAAATCCATACAGAAATGAGAGTGGCGCTTCCCGACTTCTTCAAGGCTTT 720
Db 700 AACTGAGAAATCCATACAGAAATGAGAGTGGCGCTTCCCGACTTCTTCAAGGCTTT 759
QY 721 TAGGAGTGTCTGCGGCTGCGAGGTTCCAGGAGATCAAGGATTTCAAGATTTCTAC 780
Db 760 TAGGAGTGTCTGCGGCTGCGAGGTTCCAGGAGATCAAGGATTTCAAGATTTCTAC 819
QY 781 CTTTCCATAGCAGATCATTATAGAGTTCCTGGAATGCAAAATACAGTGTGAAGAGAAC 840
820 CTTTCCATAGCAGATCATTATAGAGTTCCTGGAATGCAAAATACAGTGTGAAGAGAAC 879
841 CTCACCCAGTATAGCAGGCTATCCGGTTGAGAAATTTGCTGCTACCATGATCATTCAT 900
880 CTCACCCAGTATAGCAGGCTATCCGGTTGAGAAATTTGCTGCTACCATGATCATTCAT 939
901 TTGCAGTTTGCCTATTATAAGTTGAACGACCTGAAGAAATCGACCCCTCTGCGAGTCAG 960
940 TTGCAGTTTGCCTATTATAAGTTGAACGACCTGAAGAAATCGACCCCTCTGCGAGTCAG 999
961 TATCTGCTCTTTGATCAGATGACAGGTCATGACGAGAACCTGCTGTTATTTACAGTAC 1020
1000 TATCTGCTCTTTGATCAGATGACAGGTCATGACGAGAACCTGCTGTTATTTACAGTAC 1059
1021 CACAGGACACCTTTGGGGCTCTCGGATGACACTTTCAGCCCGCAGACCTGAAGCAGTTCAG 1080
1060 CACAGGACACCTTTGGGGCTCTCGGATGACACTTTCAGCCCGCAGACCTGAAGCAGTTCAG 1119
1081 TTCTTTAATGTGACCACTCCAGAGGAGCTGTATGACTTTTCTAGGAAATATAATG 1140
1120 TTCTTTAATGTGACCACTCCAGAGGAGCTGTATGACTTTTCTAGGAAATATAATG 1179
1141 GATGATGATGAGGAGAGTGTGGAAATATGTGGATGACCTCTTGAACCTGGAGGAGACC 1200
1180 GATGATGATGAGGAGAGTGTGGAAATATGTGGATGACCTCTTGAACCTGGAGGAGACC 1239
1201 AGC 1203
1240 AGC 1242

RESULT 6
AK075401 2010 bp mRNA linear PRI 03-SEP-2002
LOCUS Homo sapiens cDNA FSEC091 fis, clone PLACE1001683, highly similar
DEFINITION to Homo sapiens mRNA for cartilage-associated protein.
ACCESSION AK075401
VERSION AK075401.1 GI:22761465
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 Ota, T., Nishikawa, T., Suzuki, Y., Kawai-Hio, Y., Hayashi, K.,
AUTHORS Ishii, S., Saito, K., Yamamoto, J., Wakamatsu, A., Nagai, T.,
Nakamura, Y., Sato, H., Negahari, K., Sugano, S. and Isogai, T.
HRI human cDNA sequencing project; cDNA 5' - & 3' -end one pass
sequencing, clone selection and full insert sequencing; Helix
Research Institute (supported by Japan Key Technology Center etc.);
cDNA library construction; Institute of Medical Science, University
of Tokyo, Laboratory of Genome Structure, Human Genome Center.
2 Unpublished
TITLES (bases 1 to 2010)
JOURNAL Isogai, T. and Yamamoto, J.
REFERENCE Direct Submission
JOURNAL Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986
HRI human cDNA sequencing project; cDNA 5' - & 3' -end one pass
sequencing, clone selection and full insert sequencing; Helix
Research Institute (supported by Japan Key Technology Center etc.);
cDNA library construction; Institute of Medical Science, University
of Tokyo, Laboratory of Genome Structure, Human Genome Center.
FEATURES
  source
    Location/Qualifiers
      1..2010
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="PLACE1001683"
        /issue_type="placenta"
        /clone_lib="PLACE1"
        /notes="cloning vector: pME18SFL3"
      misc_difference 728^729
        /notes="compared to AK001634 and AJ006470"
        /replaces="c"

BASE COUNT 416 a 606 c 485 g 503 t
```

Query Match	98.7%	Score 1187	DB 9	Length 2010
Best Local Similarity	99.6%	Pred. No. 7e-180		
Matches 1198	Conservative	2	Mismatches 2	Indels 1
QY	1	ATGAGACGGGGCGCGCGGGGGCGCGCGCTGCTAGCGTGTGCGTGCCTGCGCTTCGCG		60
DB	96	ATGAGACGGGGCGCGGGGGGGCGCGCGCTGCTAGCGTGTGCGTGCCTGCGCTTCGCG		155
QY	61	CTGCGCGCGGGCGCGCGCCCAATACAAACGCTTACAGCTTCCGACAGCTTCCACGGGACAG		120
DB	156	CTGCGCGCGGGCGCGCGCCCAATACAAACGCTTACAGCTTCCGACAGCTTCCACGGGACAG		215
QY	121	CTGATGCGCGCTGAGTTCGGCTACGGGACGCGCTGACAAAGTACAGCGCGGACACTGG		180
DB	216	CTGATGCGCGCTGAGTTCGGCTACGGGACGCGCTGACAAAGTACAGCGCGGACACTGG		275
QY	181	GCGGAGAGCTKGGCTACTCTGGAGATCAGCTCGCGCTGCACCGCTTGTGTGCGCACAGC		240
DB	276	GCGGAGAGCGTGGCTACTCTGGAGATCAGCTCGCGCTGCACCGCTTGTGTGCGCACAGC		335
QY	241	GAGGCTTCTGGCACCGGAACTGACGCGCGCGCGCGAGCCCGCCCGCGCGCTTC		300
DB	336	GAGGCTTCTGGCACCGGAACTGACGCGCGCGCGCGAGCCCGCCCGCGCGCTTC		395
QY	301	GCCAGCTATCCGAGCTCGGCTCTTTCGGGGGCTGCTGCGCGCGCGCACTGCTCTCAAG		360
DB	396	GCCAGCTATCCGAGCTCGGCTCTTTCGGGGGCTGCTGCGCGCGCGCACTGCTCTCAAG		455
QY	361	CGCTGCACGAGGGCTGCGAGCTTTCGGGACGCTCCAGCCGCGCGAGTGTGGCG		420
DB	456	CGCTGCACGAGGGCTGCGAGCTTTCGGGACGCTCCAGCCGCGCGAGTGTGGCG		515
QY	421	GACTTCCAGCGCGGAGCCCTCAAGTTCCTGCAGTTTCGCTTACGTCGCAATAT		480
DB	516	GACTTCCAGCGCGGAGCCCTCAAGTTCCTGCAGTTTCGCTTACGTCGCAATAT		575
QY	481	CTCCCAAGGCATCGCGCTGCTCACACCTTTCCTACTGAAGCATCCTGATGACGAATG		540
DB	576	CTCCCAAGGCATCGCGCTGCTCACACCTTTCCTACTGAAGCATCCTGATGACGAATG		635
QY	541	ATGAAGAGAACATGGCATTATTAAGAGCTCGCTGCTGGTCCGAGGATACATTTAAGAC		600
DB	636	ATGAAGAGAACATGGCATTATTAAGAGCTCGCTGCTGGTCCGAGGATACATTTAAGAC		695
QY	601	CTGGAAACCAAGTCATATGAAGCTCTTTCATCCGAGCAGTGGGSCATACAAAGGTCAG		660
DB	696	CTGGAAACCAAGTCATATGAAGCTCTTTCATCCGAGCAGTGGGSCATACAAAGGTCAG		754
QY	661	AACCTGGAAACATCCATCAGACATGAGCTGGCCCTTCCGACCTTCTCAAGACCTTT		720
DB	755	AACCTGGAAACATCCATCAGACATGAGCTGGCCCTTCCGACCTTCTCAAGACCTTT		814
QY	721	TACGAGTGTCTCGAGCGCTCGAGGGTTCAGAGGATCAAGGATTCGAGGATTTCTAC		780
DB	815	TACGAGTGTCTCGAGCGCTCGAGGGTTCAGAGGATCAAGGATTTCTAC		874
QY	781	CTTTCCATAGCAGATCATTTATGTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAGAC		840
DB	875	CTTTCCATAGCAGATCATTTATGTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAGAC		934
QY	841	CTCACCCCGATTATAGAGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTTAC		900
DB	935	CTCACCCCGATTATAGAGGCTATCCGGTTGAGAAATTTGTGGCTACCATGTATCATTTAC		994
QY	901	TTGCAGTTTGCTATTATAAGTTGAACGACCTCAAGAAATGACGCCCTGTGCGAGTCAGC		960
DB	995	TTGCAGTTTGCTATTATAAGTTGAACGACCTCAAGAAATGACGCCCTGTGCGAGTCAGC		1054
QY	961	TATCTGCTCTTTGATCAGAAATGACAGGTATGACAGAACTGGGTATTACAGTAC		1020
DB	1055	TATCTGCTCTTTGATCAGAAATGACAGGTATGACAGAACTGGGTATTACAGTAC		1114



ALPDPKAFVECLAAECGSEIRIKDFDPYLSIADHVYVELECKIQEENLTPVIGYP  
VEKFAVMTHYLOFYKLNLDKNAAPCAVSYLLFPQNDKMQOINLVYGYHRDTWGL  
SDEHFOPRPEAFOFFNYNTTLQLEKLYDPKAKENIMDDGEVEVYVDLLEETS"  
413 a 556 c 467 g 476 t  
BASE COUNT  
ORIGIN

Query Match	92.4%;	Score 1111.4;	DB 9;	Length 1922;
Best Local Similarity	96.5%;	Pred. No. 8.2e-168;		
Matches 1161;	Conservative	1;	Mismatches 2;	Indels 39; Gaps 1;

QY	1	ATGAGACGGGGCGCGGGGGCGCGCGCTGCTAGCGTGTGTGCGTGGCGCTGCGCG	60
Db	40	ATGAGACGGGGCGCGGGGGCGCGCGCTGCTAGCGTGTGTGCGTGGCGCTGCGCG	99
QY	61	CTGCGCGCGGGCGCGCCCAATACGAACGTACAGCTTCGCGAGCTTCCACGGGACGAG	120
Db	100	CTGCGCGCGGGCGCGCCCAATACGAACGTACAGCTTCGCGAGCTTCCACGGGACGAG	159
QY	121	CTGATGCGGCTCGAGTCGGCTACCGGACCGGCTGGACAAGTACAGCGCGGACGACTGG	180
Db	160	CTGATGCGGCTCGAGTCGGCTACCGGACCGGCTGGACAAGTACAGCGCGGACGACTGG	219
QY	181	GCCGAGAGCTKGGCTACTCGAGATCAGCGTGGCTGCACCGCTTGTCGGCGGACGAC	240
Db	220	GCCGAGAGCG-----TGCGCGACGAC	240
QY	241	GAGGCTTCTGCAACCGCAACTGCAGCGCGCGCGCGAGCCCGCGCGCGCGCTC	300
Db	241	GAGGCTTCTGCAACCGCAACTGCAGCGCGCGCGCGAGCCCGCGCGCGCGCTC	300
QY	301	GCCAGCTATCCGAGCTCGGCTCTTGGGGGGCTGTCGCGCGCGCGGACCTGCTCAG	360
Db	301	GCCAGCTATCCGAGCTCGGCTCTTGGGGGGCTGTCGCGCGCGCGGACCTGCTCAG	360
QY	361	CGCTGAAGCAGGGCTGCCAGCTTCGCGAGCTCCGACGCCAGCGCGGAGTGTGGCG	420
Db	361	CGCTGAAGCAGGGCTGCCAGCTTCGCGAGCTCCGCGAGTCCGCGCGCGGAGTGTGGCG	420
QY	421	GACTTCCAGCGCGGAGCGCTTCAAGTTCTCGAGTTCGTTACTTCAAGGCAATAAT	480
Db	421	GACTTCCAGCGCGGAGCGCTTCAAGTTCTCGAGTTCGTTACTTCAAGGCAATAAT	480
QY	481	CTCCCCAAGCCATCGCGCTGCTCACACCTTTCTACTGAAGCATCTCTGACGAAATG	540
Db	481	CTCCCCAAGCCATCGCGCTGCTCACACCTTTCTACTGAAGCATCTCTGAGCAATG	540
QY	541	ATGAGAGGACATGGCATATTAAGAGCGCTGCTGTCGCGAGACTACATTAAAGAC	600
Db	541	ATGAGAGGACATGGCATATTAAGAGCGCTGCTGTCGCGAGACTACATTAAAGAC	600
QY	601	CTGGAACCAAGTCATATGAAGCGCTGTTTCATCGAGCAGTGGGGCATACACGCTGAG	660
Db	601	CTGGAACCAAGTCATATGAAGCGCTGTTTCATCGAGCAGTGGGGCATACACGCTGAG	660
QY	661	AACTGGAGAACTCCATCACAGCATGGAGCTGGCCCTTCCGAGCTTCTTCAAGCGCTTT	720
Db	661	AACTGGAGAACTCCATCACAGCATGGAGCTGGCCCTTCCGAGCTTCTTCAAGCGCTTT	720
QY	721	TACGAGTGTCTCGCAGCGCTCGAGGGTTCAGGGAGATCAAGGACTTCAAGGATTTCTAC	780
Db	721	TACGAGTGTCTCGCAGCGCTCGAGGGTTCAGGGAGATCAAGGACTTCTTAC	780
QY	781	CTTTTCCATAGCAGATCATATGTAGAAGTTCGGAATGCAAAATACAGTGTGAAGAGAC	840
Db	781	CTTTTCCATAGCAGATCATATGTAGAAGTTCGGAATGCAAAATACAGTGTGAAGAGAC	840
QY	841	CTCACCCAGGTTATAGGAGGCTATCCGGTTGAGAAATTTGTGCTACCAATGATCATTAC	900
Db	841	CTCACCCAGGTTATAGGAGGCTATCCGGTTGAGAAATTTGTGCTACCAATGATCATTAC	900
QY	901	TTGCGATTTGCCATATTAAGATTGAACGACTGAAGATGCAGCCCGCTGTCGAGTCAGC	960
Db	901	TTGCGATTTGCCATATTAAGATTGAACGACTGAAGATGCAGCCCGCTGTCGAGTCAGC	960

Qy	961	TATCTGCTCTTTGATCAGAAATGACAGGTCATCCAGCAGAACTGGTGTTATTACCACTAC	1020
Db	961	TATCTGCTCTTTGATCAGAAATGACAGGTCATCCAGCAGAACTGGTGTTATTACCACTAC	1020
Qy	1021	CACAGGACACTTTGGGGCCTCTCGAATGAGCACTTCCAGCCCAAGACCTGTAAGCAGTTTCAG	1080
Db	1021	CACAGGACACGTGGGGCCTCTCAGATGAGCACTTCCAGCCCAAGACCTGTAAGCAGTTTCAG	1080
Qy	1081	TTCTTTAAATGTGACCACTCCAGAAAGGAGCTATGACTTTGCTAAGGAAAAATAAATG	1140
Db	1081	TTCTTTAAATGTGACCACTCCAGAAAGGAGCTATGACTTTGCTAAGGAAAAATAAATG	1140
Qy	1141	GATGATGATCAGGAGGAAGTTGTGGAATATGTGGATGACCTCTTGGAACTGGAGGAGACC	1200
Db	1141	GATGATGATCAGGAGGAAGTTGTGGAATATGTGGATGACCTCTTGGAACTGGAGGAGACC	1200
Qy	1201	AGC 1203	
Db	1201	AGC 1203	

RESULT 8	
AK092990	
LOCUS	1903 bp mRNA linear PRI 15-JUL-2002
DEFINITION	Homo sapiens cDNA FLJ35671 fis, clone SPLEN2018180, highly similar to Homo sapiens mRNA for cartilage-associated protein.
ACCESSION	AK092990
VERSION	AK092990.1 GI:21751708
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE  
AUTHORS  
Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S.,  
Fukuzuma, Y., Fujimori, Y., Koniyama, M., Sugiyama, T., Irie, R.,  
Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J.,  
Isono, Y., Kawai, H., Ota, Y., Saito, K., Nishikawa, T., Kimura, K.,  
Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H.,  
Kande, K., Wagatsuma, M., Murakawa, K., Kanehori, K.,  
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,  
Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEDO human CDNA sequencing project  
TITLE  
1. mammalia; eutheria; primates; catarrhini; hominidae; homo.

Title  
 NEDO human cDNA sequencing project  
 Reference  
 2 (bases 1 to 1903)  
 Authors  
 Isogai, T. and Yamamoto, J.  
 Title  
 Direct Submission  
 Journal  
 Submitted (04-JUN-2003) Takao Isogai  
 Project (UPI Team) : 2-6-7

DATE	JOURNAL	COMMENT
Direct submission Submitted (04-JUL-2002)	Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazuea-kamatori.kisara@riken.ac.jp, Chiba 252-0812, Japan (E-mail:genomics@kisara.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; CDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.	

```

FEATURES
source
1..1903
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SPLEN2018180"
/tissue_type="Spleen"
/clone_lib="SPLEN2"
/clone="cloning vector: pME18SFL3"
421 a 553 c 460 g 469 t
BASE COUNT

```

Query Match 89.5%; Score 1076.6; DB 9; Length 1903;  
Best Local Similarity 95.4%; Pred. No. 2.9e-162;



13	CGCGGGGGCGCGGGCGCTGGCTAGCGCTGCTGTGGCTGCTGTGGCTGCGCGCGGG	72
60	CGAGCGCTACGGCGCGGTGTGCTGTGCTGTGCTGCGGTGCGCGCGACCCCGGG	119
73	CGCGCCCAATACGAACGCTACAGCTTCGCGAGCTTCCACGGGACAGCTGATCGCGCTC	132
120	CGGGGAGTACGAGCGCTACGCTTCCGAGCTTCCCGGGACAGCTGATGCGCGCTC	179
133	GAGTCGGCTTACCGGCACGCGCTTGGACAAGTACAGCGGCGAGCACCTGGGCCGAGAGCSTK	192
180	GAGTCGGCTTACCGCACCGGCTCGACCATGACAGCGGCGAGCACTGGGCCGAGAGCGTG	239
193	GGCTACTGGAGATCAGCCTGCGGCTGCACCGCTTGGTCGCAAGAGGCGCTTCTGC	252
240	GGCTACTGGAGTGAGCCTTGGGCTGCACCGTCTGTCGCGACACGAGGCGCTTCTGC	299
253	CACCGCAACTGCAAGCCCGCGCGCAGCCGAGCCCGCGCGGCTCGCCAGCTATCCG	312
300	CACCGCAACTGCAAGCGGCGACCGCAGACCCGCGCCCGCGCGCCGCGCCAGCCAGCC	359
313	GAGTCGCGCTTCTCGGGGGCGCTGTGTCGCGCGCGCACTGCCCTCAAGCGCTGCAAGCAG	372
360	GAATACGCGCTTCTTGGCAGCGTGTGTCGCGCGCGCAGTGCTCTCAAGCGCTGCAAGCAG	419
373	GGCCTGCCAGCCTTCGCGCAGTCCGACGCCAGCCGCGAGGTTGTCGCGACTTCCAGCGC	432
420	GGCTGCGCGCGCTTCGCGCAGTGCAGACCCAGCGCTCAGTGTGTCGCGCACTTTCAGCAG	479

[illegible]





Qy 431 GCCGCGAGCCTACAGTTCTCCTGAGTTCCTTACTTCAAGCAATATCTCCCAAG 490  
Db 443 GCCGAGAGCCTCAAGTACTCGAGTCCCTACTTCAAGGTAATATCCCTTCCAAAG 502  
Qy 491 CCATCGCGCTGTCTACACCTTTCTACTGAAGCATCTCTGATGACGAAATGATGAAGGA 550  
Db 503 CTATTGCGAGCCTCACACATTTCTTCTGAAACATCCAGATGATGAATGATGAAGGA 562  
Qy 551 ACATGGCATATTATAGAGCCTGCTGTGTCGCCAGGACTACATTAAGACCTTGGAAACCA 610  
Db 563 ATATGGCCTACTATAAGAGCATACCTGATCTGAGGAGCACATTAAGACTTGGAGACAA 622  
Qy 611 AGTCATATGAAGCCTTTCATCGAGAGCTGCGGCATACACAGTGTGAGAACTGGAGAA 670  
Db 623 AGCCTTACGAGATCTTTTGTGAGGCTGTGAGACATACATGTGTGACACTGGAGAA 682  
Qy 671 CATCCATCAGACATGAGCTGCGCCTTCCGACTTCTCAAGCCTTTTACGAGTGTG 730  
Db 683 CATCCATCTCAGATATGGAATGCTTCTCTGATTTCTCAAAACCTACGATGACTGTA 742  
Qy 731 TCGCAGCTCGAGGTTCCAGGAGATCAAGACTTCAGGATTTCTACCTTCCATAG 790  
Db 743 TAGCAGCTGTGAGGCTCCGAGAGATTAAGATTTTAAGACTTCTATCTCTCTATG 802  
Qy 791 CAGATCATTATGATGAGTTCTCGAATGCAAAATACAGTGTGA 833  
Db 803 CAGATCATTATGATGAGTCTTGCATGCAAAAGTCAGTTCGA 845

## RESULT 14

AC112211  
LOCUS Homo sapiens chromosome 3 clone RP11-627J17, complete sequence.  
DEFINITION AC112211  
ACCESSION AC112211  
VERSION AC112211.2 GI:21166208  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 182509)  
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z., Saenphimmachak,C., Phelps,K.A., Buckley,D., Raymond,C. and Haugen,E.D.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 182509)  
AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.  
TITLE Direct Submission  
JOURNAL Submitted (20-FEB-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

REFERENCE 3 (bases 1 to 182509)  
AUTHORS Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z., Saenphimmachak,C., Phelps,K.A., Buckley,D., Raymond,C. and Haugen,E.D.

TITLE Direct Submission  
JOURNAL Submitted (24-MAY-2002) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
COMMENT On May 24, 2002 this sequence version replaced gi:18767508.

Center: University of Washington Genome Center  
Center Code: UWGC  
Web site: <http://www.genome.washington.edu>  
Contact: [uwgctgs@u.washington.edu](mailto:uwgctgs@u.washington.edu)

Center project name: chr-3  
Center clone name: RP11-627J17 (bc0502)  
----- Summary Statistics

Sequencing vector: plasmid; 100% of reads  
Chemistry: Dye-terminator ET; 50% of reads  
Chemistry: Dye-terminator Big Dye; 50% of reads  
Assembly program: Phrap; version 0.990319

Consensus quality: 182361 bases at least Q40  
Consensus quality: 182497 bases at least Q30  
Consensus quality: 182509 bases at least Q20  
Insert size: 182509; sum-of-contigs  
Quality coverage: 7.9x in Q20 bases; sum-of-contigs

## Overlapping Sequences:

5: Mapping in progress  
3: Mapping in progress

## Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

## ECORI

SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	BglII
8696	8777	4710	4671	2282	2251	
6	<800	6382	6514	2067	2068	
8311	8264	512	<800	5626	5794	
3794	3847	449	<800	4289	4117	
517	<800	1417	1461	4382	4412	
5654	5717	5366	5126	4594	4580	
3530	3552	590	<800	10637	10624	
2635	2737	3606	3651	4577	4580	
3532	3552	274	<800	246	<800	
12069	11650	4570	4539	933	948	
8046	8264	9049	8995	15529	15666	
2974	3009	4374	4320	184	<800	
3451	3552	4978	5126	4296	4283	
2090	2126	11396	11164	2736	2818	





Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [gcaps-x@mail.nih.gov](mailto:gcaps-x@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: [amadan@systemsbiology.org](mailto:amadan@systemsbiology.org)

Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia

Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 7 Row: f Column: 22

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 5454037.

Location/Qualifiers

1. 1889

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="LocusID:10609"

/db\_xref="taxon:9606"

/clone="MGC:1322 IMAGE:3510406"

/tissue\_type="Brain, neuroblastoma"

/clone\_lib="NIH\_MGC\_19"

/lab\_host="DH10B-R"

/note="Vector: pOTE7"

47. 1360

/codon\_start=1

/product="nucleolar autoantigen (55kd) similar to rat

synaptonemal complex protein"

/protein\_id="AAH01047.1"

/db\_xref="GI:12654443"

/translation="MARVAVGLLLWLLSAGAYEKYSFRGPPEDLMLPAAAYGHAL

EQEGESRESARYLEALRLHRLDSEAFCHANCSPAPAAKPDGGRADSEWACE

LLRFLRERACLRCKRTLPFQVPPRQLLRDFOSRPLPYOYLHVALFKANLXG

AVAAAYTFLQRPNPKELTKYLNLYQGMVLDAESLTDLEAQYEAFLRAVKLNLS

DFRSTEDMERALSVELVAFARCLAGCGAHEVDKDFYPALADLFAESLQCKVDE

ANLTNPGYFVFXVATWYHYLOFAYKLNLDVROAARSAASYMLFDPKDSVMQNLV

YRFRHAKGLSEEDQPREAMLYHNTAEIRLELFTHMVLSQSDDEMELEETPEPL

EPEDALSAPFEGEDYBEGMYADWQBPDKAGDAEAEPEPELA"

399 a 570 c 611 g 309 t

## CDS

BASE COUNT

ORIGIN

Query Match 38.8%; Score 467; DB 9; Length 1889;  
Best Local Similarity 64.1%; Pred. No. 5.7e-65;  
Matches 755; Conservative 2; Mismatches 402; Indels 18; Gaps 3;  
QY 8 CCGGGCGCGGGGGCGGGCGGCTGTAGCGCTGCTGTGCGCTGCGGCTGCGCG 67  
Db 30 CCGGGCGCGGGCGGCGGCTGCTGGTGGCTGCGGGCTGCTGTGCTGTGGCA 89  
QY 68 CCGGGCGCGGCAATACGAACCTACAGCTTCGGCAGCTTCCACGGGACAGCTGATGC 127  
Db 90 GCGCGGGGCGGCGAGTACGAGAGTACAGCTTCGGGGCTTCCGCGCGGAGGACCTGATGC 149  
QY 128 CGCTGAGTCCGCGCTACCGGCGCGGCTGCGACAGTACAGCGGCGAGCACTGGGCGGAGA 187  
Db 150 CGCTGCGCGGCGGTACGGGCGAGCTGTGGAGCACTACAGGGAGAGCTGGCGCGAGA 209  
QY 188 GSGTGGCTACTGAGATACAGCTTGGCTGCACTGCTGCTGCGGACAGCGAGGCT 247  
Db 210 GCGCGGCTACCTGGAGGGGGGCTGCGGCTGCGGCTGCGGCTGCGGCGAGGCGCT 269  
QY 248 TTGCGCAGCGCACTGCACG-----CGCGCGCGAGCGCGAGCGCGCGCGCGCC 298  
Db 270 TTGCGCAGCGCACTGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 329  
QY 299 TCGCAGCTATC-----CGAGCTGCGGCTTTCGGGGGCGCTGCTGCGCGCGCGCACT 352

Search completed: January 30, 2004, 14:49:17

Job time : 4386.75 secs

Db 330 GCGCAGCAGTGGGCGCTGCGAGCTGGGCGCTCTTCGGCGCGGCTCTCGAGCGAGCGCGCT 389  
QY 353 GCCTCAAGCGCTGCAAGCAGGCGCTGCGAGCTTCCGCCAGTCCAGCGCCAGCGCGCGAGG 412  
Db 390 GCTTGGCGGCTGCAAGCGGAGCGCTGCCGCTTCCAGGTGCCCTACCGCGCGGCGAGC 449  
QY 413 TGCTGGGCGACTTCCAGCGCGGAGCGCTTCAAGTTCCTGCGAGTTCGCTTACTTCAAGG 472  
Db 450 TGCTGCGTGACTTCCAGAGCGCGCTGCGCTACAGTACCTGCACTACGCGCTGTCAAGG 509  
QY 473 CAATATATCTCCCCAAAGCCATGCGCGCTGCTACACTTCTTACTGAAGCATCTGATG 532  
Db 510 CTAACCGGCTGGAGAGCGGCTGGCGCGGCGCTACACCTTCTCCAGAGAAACCGGAGC 569  
QY 533 ACGAATATGATGAAGAGAAACATGCGCATATATATAA---GAGCCTGCTGCTGGCGAGCACT 589  
Db 570 ACGAGCTGACCGCGCAAGTATCTCAACTACTATCAGGSGATGCTGGAGCTGCCGACGAGT 629  
QY 590 ACATTAAAGACCTGGAAACCAAGTCATATGAAGCCCTGTTTATCCGAGCATGCGGGCAT 649  
Db 630 CCCTCACGCGACCTAGAGGCGCGCGCTACAGGCGCGCTGTTCCTCCGGGCTGTGAAGCTCT 689  
QY 650 ACMAACGCTGAGAACTGGAGAAACATCCATCACAGACATGGAGCTGGCGCTTCCCGACTTCT 709  
Db 690 ACACACGCGGGATTTCCGCGAGCAGCAGGAGGACATGGAGCGGCGCTTGTGAGATACC 749  
QY 710 TCAAAGCCTTTACGAGTGTCTCGAGCCTGCGAGGTTCCAGGGAGATCAAGGACTTCA 769  
Db 750 TGGCAGTCTTTGGCGCGGCTGCTGGCGGCTGTGAAGGCGCGCATGAGCAGGTGCACTTCA 809  
QY 770 AGGATTTCTACCTTCCATAGCAGATCATTTATGAGTTCGGAATGCAAAATACAGT 829  
Db 810 AGGCTTCTACCGCGCATAGCAGATCTCTTTCAGAGTCCCTGCGAGTCAAGGTGAGT 869  
QY 830 GTGAGAGAACCTCACCGCGCTTATAGGAGCTATCCGGTTGAGAAATTTTGGCTACCA 889  
Db 870 GTGAGGCAATTTGACCCCCCAATGTGGTGGCTACTTGTGGGACAAAGTTCGTGGCCACCA 929  
QY 890 TGTATCATTTACTTGCAGTTTGCCTATTATAAGTTGAACGACCTGAAGATGCAAGCCCT 949  
Db 930 TGTACCACTACCTGCAAGTTTGCCTCTATAAGTTGAATGATGTGCGCCAGGCTGCGCGCA 989  
QY 950 GTGCGTCACTATCTGCTCTTTGATCAGAAATGACAAGGTTCATGCAAGCAAACTGCTGT 1009  
Db 990 GCGCGCGCAGCTACATGCTCTTCGACCCCAAGGACAGCGCTCATGCAAGCAAACTGCTGT 1049  
QY 1010 ATTACAGTACCAAGGACACTTGGGGCGCTCTCGGATGAGCACTTCCAGCGCCAGACCTG 1069  
Db 1050 ATTACCGGTTCCACCGGCTGCTGGGGCGCTTGGAGAGGAGGACTTCCAGCGCGCGGAG 1109  
QY 1070 AAGCAGTTCAAGTTCTTTAATGTGACCACTCCAGAAAGAGCTGTATGACTTGTCTAAGG 1129  
Db 1110 AGGCCATGCTTACCAACACACCGAGCCGCGAGCTGCGGAGCTGTGGAGTTTCAACCA 1169  
QY 1130 AAATATATGAGATGATGAGGGGAGAGTGTGGA 1166  
Db 1170 TGTACCTGCAGTCAGATGATGATGAGTGGAGCTGGAGGA 1206

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 07:10:40 ; Search time 14098.5 Seconds  
(without alignments)  
11232.506 Million cell updates/sec

Title: US-09-729-674-1

Perfect score: 3871  
Sequence: 1 ttctcttccctccctttt.....aataaaaaaaaaaaaaaa 3871

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813396 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_on.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_scs.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3870.2	100.0	3871	6	BD140451 Secreted
2	2619.6	67.7	182509	9	AC112211 Homo sapi
3	2605.2	67.3	166485	9	AC123900 Homo sapi
4	2237	57.8	2307	9	AJ006470 Homo sapi
5	1980.4	50.6	2010	9	AK075401 Homo sapi
6	1955.4	50.5	1961	6	BD156517 Primer fo
7	1955.4	50.5	1961	9	AK001634 Homo sapi
8	1869.4	48.3	1922	9	AK091772 Homo sapi
9	1798.8	46.5	1903	9	AK092990 Homo sapi
10	1656	42.8	1810	9	AK097650 Homo sapi
11	1370.4	35.4	1439	9	EC008745 Homo sapi
12	944	24.4	1685	10	MMU64649 Mus muscu
13	940	24.3	1663	10	BC049890 Mus muscu
14	695.6	18.0	818	6	BD148800 Primer fo
15	510.2	13.2	64706	2	AC136358 Homo sapi
16	497.6	12.9	1665	5	GGCARPR X97607 G. gallus
17	467	12.1	1889	9	BC001047 Homo sapi
18	467	12.1	2037	9	BC007942 Homo sapi
19	467	12.1	2443	6	AX714040 Sequence
20	467	12.1	2443	9	AK056085 Homo sapi
21	466.6	12.1	1745	9	BC011701 Homo sapi
22	465.8	12.0	2347	6	AX335490 Sequence
23	465.8	12.0	2347	9	HSU47621 Homo sapien
24	465.4	12.0	2079	9	HSU250583
25	428.5	11.1	1903	10	BC031856
26	403	10.4	1407	10	RNSC65MR
27	392.4	10.1	394	6	AX408806
28	369.4	9.5	400	11	G27784 human STS S
29	355.6	9.2	407	6	AX150162 Sequence
30	352.2	9.1	226615	2	AC122652 Rattus no
31	340	8.8	351	6	BD153832 Primer fo
32	313.6	8.1	330	6	AX150140 Sequence
33	283.4	6.9	274	6	AX333823 Sequence
34	258.6	6.7	279	6	AX261032 Sequence
35	252.2	6.5	157326	9	AC008810 Homo sapi
36	250.6	6.5	257967	2	AL365337 Mus muscu
37	249.6	6.4	164217	9	AC093264 Homo sapi
38	249	6.4	200240	2	AC011618
39	247.8	6.4	157375	9	AC025207 Homo sapi
40	247.8	6.4	178420	9	AP003548 Homo sapi
41	247.8	6.4	199569	9	AC016250 Homo sapi
42	247.8	6.4	211076	2	AC012214 Homo sapi
43	247.4	6.4	189533	9	AC025678 Homo sapi
44	246.6	6.4	115136	9	AL390994 Human DNA
45	245.8	6.3	169089	9	AC008751 Homo sapi

ALIGNMENTS

RESULT 1  
BD140451  
LOCUS BD140451 3871 bp DNA linear PAT 18-SEP-2002  
DEFINITION Secreted proteins and polynucleotides encoding them.  
ACCESSION BD140451  
VERSION BD140451.1 GI:23235396  
KEYWORDS JP 2002506611-A/1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 3871)  
AUTHORS Jacobs,K., Mccoy,J.M., Lavallie,E.R., Racie,L.A.C., Evans,C.,  
Merberg,D., Treacy,M., Agostino,M.J., II,R.J.S., Wong,G.G.,  
Clark,H.F. and Fichtel,K.



1681 DB CAGGCTGGTCTCGAACTCTTGACCTTCAGATGATCAATCTGGCCCTCCCAAGTGC 1740  
1741 QY TGGGAATTA CAGCGTGAGCAACATGCCCCGGCTCTTTCTACCTTTTACACCTGCTTCT 1800  
1741 DB TGGGAATTA CAGCGTGAGCAACATGCCCCGGCTCTTTCTACCTTTTACACCTGCTTCT 1800  
1801 QY TATCTCTACATCTGTTTTCACACCTTCATCCCTGCTCTTCTCATGTTTACACCTGCTTCT 1860  
1801 DB TATCTCTACATCTGTTTTCACACCTTCATCCCTGCTCTTCTCATGTTTACACCTGCTTCT 1860  
1861 QY CCCATGTTTCATAGCTGCTTTCTTACCAATTTGGTTTGAAGGCGAGTCTTCTCTGGCTTG 1920  
1861 DB CCCATGTTTCATAGCTGCTTTCTTACCAATTTGGTTTGAAGGCGAGTCTTCTCTGGCTTG 1920  
1921 QY TTTTCTTGTCTTCCCGAATAATCAGTATATTTTTTAAATAAGAAAACATTTCTAGAA 1980  
1921 DB TTTTCTTGTCTTCCCGAATAATCAGTATATTTTTTAAATAAGAAAACATTTCTAGAA 1980  
1981 QY GATGATAATTTGTAATACTCTTGGCTTATTTGCTTTTCCAGATTTTAGTCTCCTTTC 2040  
1981 DB GATGATAATTTGTAATACTCTTGGCTTATTTGCTTTTCCAGATTTTAGTCTCCTTTC 2040  
2041 QY TCCCATCCGGAAAGATGTTGGAAGACATAGGCTTAAATTTCTCCAGCTTCACAATGCTC 2100  
2041 DB TCCCATCCGGAAAGATGTTGGAAGACATAGGCTTAAATTTCTCCAGCTTCACAATGCTC 2100  
2101 QY TTCACCTTGGTCTGACCTTGACCAATTTAGCACCCACTGAAAACAGATTGAGTAGAGAG 2160  
2101 DB TTCACCTTGGTCTGACCTTGACCAATTTAGCACCCACTGAAAACAGATTGAGTAGAGAG 2160  
2161 QY TGTAGAGTGCAGAAATGTTGGCTTTTGGCCCCACTTTGCACTTCCAAATTTCAACGGTTGG 2220  
2161 DB TGTAGAGTGCAGAAATGTTGGCTTTTGGCCCCACTTTGCACTTCCAAATTTCAACGGTTGG 2220  
2221 QY CCGATCCCATTTGAGGACAACTGTTAGTTTAAAGTCTCCGAGTTTGGAAAAGGAAGAAGC 2280  
2221 DB CCGATCCCATTTGAGGACAACTGTTAGTTTAAAGTCTCCGAGTTTGGAAAAGGAAGAAGC 2280  
2281 QY CAGAGCTCTAGTTTTCATTCCTTTCAGTAAATATTTTATGAGTACCTTACTGCTGTC 2340  
2281 DB CAGAGCTCTAGTTTTCATTCCTTTCAGTAAATATTTTATGAGTACCTTACTGCTGTC 2340  
2341 QY TAGGCAATTTGACCTGGGAACTAGAGATCTTACAGAAATTAACAGGGAAGTTCCCTGTCT 2400  
2341 DB TAGGCAATTTGACCTGGGAACTAGAGATCTTACAGAAATTAACAGGGAAGTTCCCTGTCT 2400  
2401 QY CATGAGCTTACATTTCTACAGGGAAGAGATAGCCATACATAGGAATAAATATATAC 2460  
2401 DB CATGAGCTTACATTTCTACAGGGAAGAGATAGCCATACATAGGAATAAATATATATAC 2460  
2461 QY AAGGTATCATGTAGTGAATTTGCTGTGGAAAAATAAAGCAGGGGAGGAGTAAAGAAA 2520  
2461 DB AAGGTATCATGTAGTGAATTTGCTGTGGAAAAATAAAGCAGGGGAGGAGTAAAGAAA 2520  
2521 QY TCCTGGAGATGAGCTGAGTTTTTAAATGGGGCTCTACTGGGAATGTGAGTGGAGAGA 2580  
2521 DB TCCTGGAGATGAGCTGAGTTTTTAAATGGGGCTCTACTGGGAATGTGAGTGGAGAGA 2580  
2581 QY GACGTTAGGGAAGTGGATTCCTGGCAAGGCAATTCAGGCGAGAGAAACAAGATGTGCACTG 2640  
2581 DB GACGTTAGGGAAGTGGATTCCTGGCAAGGCAATTCAGGCGAGAGAAACAAGATGTGCACTG 2640  
2641 QY CCCAAAGTGAGAACTTGTCTCTAGTGTGAGGAAAGAGCAGGGAGACCAGCAGAGTGC 2700  
2641 DB CCCAAAGTGAGAACTTGTCTCTAGTGTGAGGAAAGAGCAGGGAGACCAGCAGAGTGC 2700  
2701 QY TGGGCGGGGTAGAAATGGAAGGAGCGGCTGGGAGGACAGTGGTGGAGGCGCTTGG 2760  
2701 DB TGGGCGGGGTAGAAATGGAAGGAGCGGCTGGGAGGACAGTGGTGGAGGCGCTTGG 2760  
2761 QY CTTCTGTAGTGGGAACCACTGGAGGGTTTGAACAGAGGAGTGCCTTGAATGAT 2820  
2761 DB CTTCTGTAGTGGGAACCACTGGAGGGTTTGAACAGAGGAGTGCCTTGAATGAT 2820

2821 QY TTATATTTTCAAGGGTCATTTAGCTGCAATATTGTGAAAACTTTTAGTGGCAAGGGC 2880  
2821 DB TTATATTTTCAAGGGTCATTTAGCTGCAATATTGTGAAAACTTTTAGTGGCAAGGGC 2880  
2881 QY AGAAGGAAGAGGAAGAACCTGTTTAGGAAGCTACTGCAAGGTTCCAGGCTTGGCCCTGGC 2940  
2881 DB AGAAGGAAGAGGAAGAACCTGTTTAGGAAGCTACTGCAAGGTTCCAGGCTTGGCCCTGGC 2940  
2941 QY CACAGCAACAGCAGTGTGCTCAATATCTAGATTTTATTTGAAAGAGCCCAATAGGATTGC 3000  
2941 DB CACAGCAACAGCAGTGTGCTCAATATCTAGATTTTATTTGAAAGAGCCCAATAGGATTGC 3000  
3001 QY TGAGAGTTTGAATGTGAGTGTAAAGAGGAAGAGTAAATGATGACATTTAAGGTTTTTG 3060  
3001 DB TGAGAGTTTGAATGTGAGTGTAAAGAGGAAGAGTAAATGATGACATTTAAGGTTTTTG 3060  
3061 QY GCCTGAATAGCAGGAAGATGGAGTTTACCAGTTACTGAAATAGGAGAGTGGCTGGGT 3120  
3061 DB GCCTGAATAGCAGGAAGATGGAGTTTACCAGTTACTGAAATAGGAGAGTGGCTGGGT 3120  
3121 QY AAGTAAAGGAATTTGGTCAAGCAGGCTCTCTGTGTTTGAATGGAGGTTCTGGCTGCA 3180  
3121 DB AAGTAAAGGAATTTGGTCAAGCAGGCTCTCTGTGTTTGAATGGAGGTTCTGGCTGCA 3180  
3181 QY AATCAAGTGGAGATTTCTCTCAGGTCAGGCTGTCAGCAGAGCTCGAGACAGGATCTGAA 3240  
3181 DB AATCAAGTGGAGATTTCTCTCAGGTCAGGCTGTCAGCAGAGCTCGAGACAGGATCTGAA 3240  
3241 QY TGCACCTTGGTTTATTTGTTGGGGTCTCTCAGAGGAACCTGTGAAAGCCTTTATCAGTC 3300  
3241 DB TGCACCTTGGTTTATTTGTTGGGGTCTCTCAGAGGAACCTGTGAAAGCCTTTATCAGTC 3300  
3301 QY ATTTATGCTGTGAGAAATTTCTCTGGAGTGTGGGTACATTTGAAGGCAAGTACTTCA 3360  
3301 DB ATTTATGCTGTGAGAAATTTCTCTGGAGTGTGGGTACATTTGAAGGCAAGTACTTCA 3360  
3361 QY GTTGGGGCAAGTCTCTGAAAAGAGGCTGTAGGCTCTGGCAGCTACCATCGTGGTAG 3420  
3361 DB GTTGGGGCAAGTCTCTGAAAAGAGGCTGTAGGCTCTGGCAGCTACCATCGTGGTAG 3420  
3421 QY TGTGTTGGGGTGGGGTCTCTGGCAGTGTGTGAAGGATCTGGCAGGCAACCACA 3480  
3421 DB TGTGTTGGGGTGGGGTCTCTGGCAGTGTGTGAAGGATCTGGCAGGCAACCACA 3480  
3481 QY GCGCCCCCTACTGAACCATCAGCATGTCTGAGGCAATTTAAAGCCATGCACTGGAGGGC 3540  
3481 DB GCGCCCCCTACTGAACCATCAGCATGTCTGAGGCAATTTAAAGCCATGCACTGGAGGGC 3540  
3541 QY CACTGAGATTGTCTGTGAGTATTACTGAGAACCAACAGAAAGCCATGGATGGAGCCC 3600  
3541 DB CACTGAGATTGTCTGTGAGTATTACTGAGAACCAACAGAAAGCCATGGATGGAGCCC 3600  
3601 QY TTGGGCTCTCTGGGAAATGGGAAATCAGCCAAAGGACTGAGAGGAGTTACCTTAAGGTC 3660  
3601 DB TTGGGCTCTCTGGGAAATGGGAAATCAGCCAAAGGACTGAGAGGAGTTACCTTAAGGTC 3660  
3661 QY ACAGAAAACCAAGAGAGTGTGTGTTCTGGAAGCTGAGCTTCTTTATTTCAACCTCATTC 3720  
3661 DB ACAGAAAACCAAGAGAGTGTGTGTTCTGGAAGCTGAGCTTCTTTATTTCAACCTCATTC 3720  
3721 QY CTTTCTCCAAATAAGCCACTTGTGTAGTTGGGCCCTCCAGGGTTGAAGCAAGAGGAGA 3780  
3721 DB CTTTCTCCAAATAAGCCACTTGTGTAGTTGGGCCCTCCAGGGTTGAAGCAAGAGGAGA 3780  
3781 QY AAGGCACAGCTTTGGGAAACAGACTTTTCTGCAATAGCTTGGGAAGGAATAAAGGA 3840  
3781 DB AAGGCACAGCTTTGGGAAACAGACTTTTCTGCAATAGCTTGGGAAGGAATAAAGGA 3840  
3841 QY TAGAGTGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3871  
3841 DB TAGAGTGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3871

RESULT 2  
 AC112211  
 LOCUS  
 DEFINITION Homo sapiens chromosome 3 clone RP11-627J17, complete sequence.  
 AC112211  
 ACCESSION  
 VERSION  
 SOURCE  
 KEYWORDS  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 182509)  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
 Saenphimmachak, C., Phelps, K.A., Buckley, D., Raymond, C. and  
 Haugen, E.D.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 182509)  
 Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.  
 Direct Submission  
 Submitted (20-FEB-2002) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 3 (bases 1 to 182509)  
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,  
 Saenphimmachak, C., Phelps, K.A., Buckley, D., Raymond, C. and  
 Haugen, E.D.  
 Direct Submission  
 Submitted (24-MAY-2002) Genome Center, University of Washington,  
 Box 352145, Seattle, WA 98195, USA  
 On May 24, 2002 this sequence version replaced gi:18767508.  
 ----- Genome Center  
 Center: University of Washington Genome Center  
 Center Code: UWGC  
 Web site: <http://www.genome.washington.edu>  
 Contact: [uwgctgs@u.washington.edu](mailto:uwgctgs@u.washington.edu)  
 ----- Project Information  
 Center project name: chr-3  
 Center clone name: RP11-627J17 (bc0502)  
 ----- Summary Statistics  
 Sequencing vector: plasmid; 100% of reads  
 Chemistry: Dye-terminator ET; 50% of reads  
 Assembly: Dye-terminator Big Dye; 50% of reads  
 Consensus quality: Phrap; version 0.990319  
 Consensus quality: 182361 bases at least Q40  
 Consensus quality: 182497 bases at least Q30  
 Consensus quality: 182509 bases at least Q20  
 Insert size: 182509; sum-of-contigs  
 Quality coverage: 7.9x in Q20 bases; sum-of-contigs  
 -----  
 Overlapping Sequences:  
 5': Mapping in progress  
 3': Mapping in progress  
 -----  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.  
 -----  
 This sequence was finished as follows unless otherwise noted:  
 all regions were either double-stranded or sequenced with an  
 alternate chemistry or covered by high quality data (i.e., Phred  
 quality >= 30); an attempt was made to resolve all sequencing  
 problems, such as compressions and repeats; all regions were  
 covered by at least one plasmid subclone or more than one M13  
 subclone; and the assembly was confirmed by restriction digest.  
 -----  
 Sequence Validation:  
 -----

This sequence has been validated by Multiple Complete Digest  
 fingerprinting. Comparison of the experimentally derived digest  
 fragments with sequence-predicted fragments is given below.  
 The electronically-digested sequence consists of both insert and  
 vector, in order to accurately represent the entire circular BAC.  
 Small fragments below a variable cutoff (approximately 400-800 bp)  
 are not resolved in the fingerprint and hence do not appear  
 in the table. There are no significant remaining discrepancies  
 between the experimental and predicted values. Uniquely ordered  
 fragments are separated by dashed lines.

ECORI	HindIII	BglII	
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
8696	8777	4710	4671
6	<800	6382	6514
8311	8264	512	<800
3794	3847	449	<800
517	<800	1417	1461
5654	5717	5366	5126
3530	3552	590	<800
2635	2737	3606	3651
3532	3552	274	<800
12069	11650	4570	4539
8046	8264	9049	8995
2974	3009	4374	4320
3451	3552	4978	5126
2090	2126	11396	11164
1581	1500	1746	1719
555	<800	1599	1541
210	<800	8101	8204
1074	1051	2394	2339
2455	2540	1944	1928
8953	8777	5118	5126
361	<800	3069	3110
3186	3259	2276	2339
19747	19152	7388	7507
1669	1668	196	<800
1367	1329	312	<800
3124	3171	2294	2339
24350	24894	5185	5126
6070	6075	714	<800
1385	1329	1355	1358
			6197
			6227

FEATURES				Location/Qualifiers			
source				1..182509			
Query Match				67.7%; Score 2619.6; DB 9; Length 182509;			
Best Local Similarity				99.8%; Pred No. 0;			
Matches 2633; Conservative				0; Mismatches 4; Indels 1; Gaps 1;			
Qy	1213	AGGGAGAAGTTGTGGAATATGTGGATGACCTCTTGGAACTGGAGGAGACAGCTAGCCCA	1272				
Db	178281	AGGGAGAAGTTGTGGAATATGTGGATGACCTCTTGGAACTGGAGGAGACAGCTAGCCCA	178340				
Qy	1273	CAGCAACCAAGAGACTTCCTCTGGCGTTCAGGAACACAGATCTTGTCTTTTCCC	1332				
Db	178341	CAGCAACCAAGAGACTTCCTCTGGCGTTCAGGAACACAGATCTTGTCTTTTCCC	178400				
Qy	1333	AACGCCAGGCTGTGTATACCTCAGAGCCTTCTCTTTACTCTCCAAAGTGAAGGGAAG	1392				
Db	178401	AACGCCAGGCTGTGTATACCTCAGAGCCTTCTCTTTACTCTCCAAAGTGAAGGGAAG	178460				

Qy	1393	CCCCGCTCTCTAACTGCAATGATGTCATCAGGGGTGAGCTGCCCTTCTCTATCTTTCACACCT	1452				
Db	178461	CCCCGCTCTCTAACTGCAATGATGTCATCAGGGGTGAGCTGCCCTTCTCTATCTTTCACACCT	178520				
Qy	1453	GCACCTCATGTTTCACACCTATCTTCTCACCTTTTTTTTGTAGATGGAGTCTCGCTCTCT	1512				
Db	178521	GCACCTCATGTTTCACACCTATCTTCTCACCTTTTTTTTGTAGATGGAGTCTCGCTCTCT	178580				
Qy	1513	TGCCAGGCTGGAGTGAATGGACAGCTTCTCAGCTCAGTCACACCTCCGCTCTTGGGTT	1572				
Db	178581	TGCCAGGCTGGAGTGAATGGACAGCTTCTCAGCTCAGTCACACCTCCGCTCTTGGGTT	178640				
Qy	1573	CAAGCAATTCGTGTCATCAGCCTCCGAGTACTTGGGATTACAGGCATGTGCACCAAG	1632				
Db	178641	CAAGCAATTCGTGTCATCAGCCTCCGAGTACTTGGGATTACAGGCATGTGCACCAAG	178700				
Qy	1633	CCCGGCTAATTTTGTATTTTGTAGTAGAGAGGGGTTTTGCCATGTTGGCCAGGCTGGTCT	1692				
Db	178701	CCCGGCTAATTTTGTATTTTGTAGTAGAGAGGGGTTTTGCCATGTTGGCCAGGCTGGTCT	178760				
Qy	1693	CGAACTCTTGACTTCAGATGATCCATCTGCCCTTGGCTCCACAGTCGTGGGATTACAGG	1752				
Db	178761	CGAACTCTTGACTTCAGATGATCCATCTGCCCTTGGCTCCACAGTCGTGGGATTACAGG	178820				
Qy	1753	CGTAGCCACCATGCCCCGCTCTTCTCACCTTTACACTGCTCTTCTTATCTCATCATC	1812				
Db	178821	CGTAGCCACCATGCCCCGCTCTTCTCACCTTTACACTGCTCTTCTTATCTCATCATC	178880				
Qy	1813	TGTTTTACACCTTCATCTCTCTCTCATGTTCCATCTTCCATCTTCTTCCCATGTTCTA	1872				
Db	178881	TGTTTTACACCTTCATCTCTCTCTCATGTTCCATCTTCCATCTTCTTCCCATGTTCTA	178940				
Qy	1873	GCTGCTCTTCTTACCAATTTGGTTGAAGGCGAGTCTTCTCTGGCTGTTTTTGTCTT	1932				
Db	178941	GCTGCTCTTCTTACCAATTTGGTTGAAGGCGAGTCTTCTCTGGCTGTTTTTGTCTT	179000				
Qy	1933	TCCAGAAATCAGTATTTTAAATAGAAACATCTCTAGAGATGATATTTCT	1992				
Db	179001	TCCAGAAATCAGTATTTTAAATAGAAACATCTCTAGAGATGATATTTCT	179060				
Qy	1993	GAACCTCTCTTGGCTTATTTGCTTTTCCAGATTTTCTCTCTCTTCTTCTTCTTCTTCT	2052				
Db	179061	GAACCTCTCTTGGCTTATTTGCTTTTCCAGATTTTCTCTCTCTTCTTCTTCTTCTTCT	179120				
Qy	2053	AAAGATGTTGGAAGACATAGGCTAAATTTCTCCAGCTCACAATGCTTCTTCTTCTTCT	2112				
Db	179121	AAAGATGTTGGAAGACATAGGCTAAATTTCTCCAGCTCACAATGCTTCTTCTTCTTCT	179180				
Qy	2113	GACTTGTACCAATTCAGACCCACTGAAAAACAAGTTGAGTAGAGAGTGTAGAGTGCAG	2172				
Db	179181	GACTTGTACCAATTCAGACCCACTGAAAAACAAGTTGAGTAGAGAGTGTAGAGTGCAG	179240				
Qy	2173	AAATGCGCTTTTGGCCCACTTTGCAATCTCCAAATTAACAAGTTGGCCGATGCCATTT	2232				
Db	179241	AAATGCGCTTTTGGCCCACTTTGCAATCTCCAAATTAACAAGTTGGCCGATGCCATTT	179300				
Qy	2233	GAGCAATGCTTAGTTTAACTCTCCGAGTTTGAAGAGGAGGAGGAGGAGGAGGAGGAGG	2292				
Db	179301	GAGCAATGCTTAGTTTAACTCTCCGAGTTTGAAGAGGAGGAGGAGGAGGAGGAGGAGG	179360				
Qy	2293	GTTCATTCATTTCTTCTAGTAAATTTTATGAGTACCTTCTGCTGCTAGGCAATGACC	2352				
Db	179361	GTTCATTCATTTCTTCTAGTAAATTTTATGAGTACCTTCTGCTGCTAGGCAATGACC	179420				
Qy	2353	TGGGAATCAGATACCTTCAGAGATACAGGGAAGTTCCCTGCTGCTAGGCAATGACC	2412				
Db	179421	TGGGAATCAGATACCTTCAGAGATACAGGGAAGTTCCCTGCTGCTAGGCAATGACC	179480				
Qy	2413	ATTCTACGGGAGAGATGACCAATACATAGGAATAAATATATACAGGATATCATGT	2472				
Db	179481	ATTCTACGGGAGAGATGACCAATACATAGGAATAAATATATACAGGATATCATGT	179540				



```
QY 2473 AGTGATTAATGCTGTGCGAGAAATAAAGCAGGGAGGAGTAAGAAATCTCTGGAGATGA 2532
Db 179541 AGTGATTAATGCTGTGCGAGAAATAAAGCAGGGAGGAGTAAGAAATCTCTGGAGATGA 179600
QY 2533 GGCTGCAGTCTTTAAATGGGGCCTCACTGGGAATGTGACCTTCAGCAGAGACGCTTAGGAA 2592
Db 179601 GGCTGCAGTCTTTAAATGGGGCCTCACTGGGAATGTGACCTTCAGCAGAGACGCTTAGGAA 179660
QY 2593 GTGGATCTGTGCAAGGCAATTCAGGCAGAGGAAACAAGATGTGCACTGCCCCAAAGTGAG 2652
Db 179661 GTGGATCTGTGCAAGGCAATTCAGGCAGAGGAAACAAGATGTGCACTGCCCCAAAGTGAG 179720
QY 2653 AACTCTCTCTACGTGTCTAGGAAGAGCAGGAGACCAACAGCAGAGTCGTGGCAGGGGTA 2712
Db 179721 AACTCTCTCTACGTGTCTAGGAAGAGCAGGAGACCAACAGCAGAGTCGTGGCAGGGGTA 179780
QY 2713 GAATGAAAGAGAGGCGCTGGGAGGACAGAGTGTGTGGAGGCCCTTGGCTTCTGCTAAGT 2772
Db 179781 GAATGAAAGAGAGGCGCTGGGAGGACAGAGTGTGTGGAGGCCCTTGGCTTCTGCTAAGT 179840
QY 2773 GAGATGGGAACCACTGGAGGGTTGAAACAGAGAGTGCCTTGAATGATTTATTTTGA 2832
Db 179841 GAGATGGGAACCACTGGAGGGTTGAAACAGAGAGTGCCTTGAATGATTTATTTTGA 179900
QY 2833 AGGTCATCTAGCTGCAATATTTGAAATACTTTAGTGGCAAGGGCAGAGGAAGAGG 2892
Db 179901 AGGTCATCTAGCTGCAATATTTGAAATACTTTAGTGGCAAGGGCAGAGGAAGAGG 179960
QY 2893 GAAGACCTGTAGGAAGTACTGCAAGGTTCCAGGCTTGGGCTGGGCCAGCAACAGC 2952
Db 179961 GAAGACCTGTAGGAAGTACTGCAAGGTTCCAGGCTTGGGCTGGGCCAGCAACAGC 180020
QY 2953 AGTGTGCAATATCTAGATTTATTTGAAAGAGCAATAGGATTTGCTGAGAGTTGAA 3012
Db 180021 AGTGTGCAATATCTAGATTTATTTGAAAGAGCAATAGGATTTGCTGAGAGTTGAA 180080
QY 3013 TGTGAGTGTAGAGAGGAGTAAATGATGAATTAAGTTTGTGCTGCAATAGCA 3072
Db 180081 TGTGAGTGTAGAGAGGAGTAAATGATGAATTAAGTTTGTGCTGCAATAGCA 180140
QY 3073 GGAAGATGGATTACCTACTGTAATAGGAGAGGATGGCTGGGTAGTAAAGTATT 3132
Db 180141 GGAAGATGGATTACCTACTGTAATAGGAGAGGATGGCTGGGTAGTAAAGTATT 180200
QY 3133 TGTGCAAGAGAGGCTGTCTGTGTGTAATGGAGGTTCTGGCTGCAATCAAGATGGA 3192
Db 180201 TGTGCAAGAGAGGCTGTCTGTGTGTAATGGAGGTTCTGGCTGCAATCAAGATGGA 180260
QY 3193 GA-TTCTCTCAGGTGAGTCTGACAGAGCTCGAGACAGGATCTGAATGCACTTGTT 3251
Db 180261 GAGTCTCTCAGGTGAGTCTGACAGAGCTCGAGACAGGATCTGAATGCACTTGTT 180320
QY 3252 TATTGTGGGGTGTCTCAGAGGAACTGTGAAAGGCTTTATCAGTCATTTATTGGCT 3311
Db 180321 TATTGTGGGGTGTCTCAGAGGAACTGTGAAAGGCTTTATCAGTCATTTATTGGCT 180380
QY 3312 GTGAGAGTCTCTGGAGTGTGGTCAATTTGAAGGCAAGTGAATTCAGTTGAGGGCAA 3371
Db 180381 GTGAGAGTCTCTGGAGTGTGGTCAATTTGAAGGCAAGTGAATTCAGTTGAGGGCAA 180440
QY 3372 GTCTCTGAAAAGAGGCTGTAGGCATCTGGCAGCTACCATGCTGGTGTAGTGTGGGG 3431
Db 180441 GTCTCTGAAAAGAGGCTGTAGGCATCTGGCAGCTACCATGCTGGTGTAGTGTGGGG 180500
QY 3432 TGGGGTCTCTGGGCACTGCTGTGTGAAGGATCTGGCAGGCAACCAAGCGCCCTTAC 3491
Db 180501 TGGGGTCTCTGGGCACTGCTGTGTGAAGGATCTGGCAGGCAACCAAGCGCCCTTAC 180560
QY 3492 TGAACCATCAGCATGTCAGTGCAATTTAAAGCCATGCACTGGAGGGCCCTCAGATTG 3551
Db 180561 TGAACCATCAGCATGTCAGTGCAATTTAAAGCCATGCACTGGAGGGCCCTCAGATTG 180620
QY 3552 TCTCTGAGTATTACTGAGAAGCAACAGAAAGAGCCATGGATGGAGGCCCTTGGGCTCTCT 3611
```

```
180621 TCCTGAGTATTACTGAGAAGCAACAGAAAGAGCCATGGATGGAGGCCCTTGGGCTCTCT 180680
QY 3612 GGGAATGGGAATCAGCCAAAGACTGGAAGAGCTTACCTTAAGGTGAGAGAAACCA 3671
Db 180681 GGGAATGGGAATCAGCCAAAGACTGGAAGAGCTTACCTTAAGGTGAGAGAAACCA 180740
QY 3672 AGAGAGTGTGGTGTCTGGAAGCTGAGCTTTCTTTTATTTCAACCTCATTTCCCTTCTCCAA 3731
Db 180741 AGAGAGTGTGGTGTCTGGAAGCTGAGCTTTCTTTTATTTCAACCTCATTTCCCTTCTCCAA 180800
QY 3732 TAAGCCACTTGTAGTGTGGCCCTCCAGGCTTGAAGCAAGAGAGAAAGCAGCG 3791
Db 180801 TAAGCCACTTGTAGTGTGGCCCTCCAGGCTTGAAGCAAGAGAGAAAGCAGCG 180860
QY 3792 TTTGGGAAACAAGACTTTTCTGCAATAGCTCGGAGGAATAAAAAGGATAGAGTGT 3849
Db 180861 TTTGGGAAACAAGACTTTTCTGCAATAGCTCGGAGGAATAAAAAGGATAGAGTGT 180918
```

RESULT 3  
AC123900  
LOCUS  
DEFINITION Homo sapiens chromosome 3 clone RP11-640D6, linear PRI 29-JUN-2002  
AC123900  
ACCESSION  
VERSION  
AC123900.2 GI:21629387  
KEYWORDS  
HTG.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 166485)  
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,  
Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.  
and Haugen,B.D.  
Direct Submission  
Unpublished  
2 (bases 1 to 166485)  
Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,B.D.  
Direct Submission  
Submitted (04-JUN-2002) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 166485)  
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,  
Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.  
and Haugen,B.D.  
Direct Submission  
Submitted (29-JUN-2002) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
On Jun 29, 2002 this sequence version replaced gi:21321810.  
----- Genome Center  
Center: University of Washington Genome Center  
Web site: <http://www.genome.washington.edu>  
Contact: [uwgctgs@u.washington.edu](mailto:uwgctgs@u.washington.edu)  
----- Project Information  
Center project name: chr-3  
Center clone name: RP11-640D6 (bc0506)  
----- Summary Statistics  
Sequencing vector: plasmid; 100% of reads  
Chemistry: Dye-terminator ET; 19% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 166343 bases at least Q40  
Consensus quality: 166451 bases at least Q30  
Consensus quality: 166480 bases at least Q20  
Insert size: 166485; sum-of-contigs  
Quality coverage: 7.5x in Q20 bases; sum-of-contigs  
-----  
Overlapping Sequences:  
5': RP11-627J17 (UWGC:bc0502) AC112211 21818-bp overlap  
3': RP11-294O18 (UWGC:bc0346) AC122176  
-----

## Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

EcoRI

BglII

HindIII

SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
8696	8737	2811	2816	1461	1410
6	<800	2067	2044	6382	6450
3530	3515	11749	11492	512	<800
5654	5722	4594	4513	449	<800
517	<800	4382	4513	8031	8087
3794	3785	4295	4513	4570	4537
8317	8204	2621	2816	274	<800
1659	1620	1611	1564	3606	3653
8141	8204	1043	1015	590	<800
646	<800	3299	3403	5372	5286
2223	2226	885	861	3079	3054
3811	3785	9258	9095	775	<800
3020	3008	282	<800	659	<800
2953	3008	2006	2044	1190	1170
3240	3226	11003	11492	1357	1328
3763	3785	8	<800	4836	4684
13647	13653	4619	4513	408	<800
3457	3515	449	<800	21	<800
379	<800	1955	2044	258	<800

FEATURES

Location/Qualifiers

10453	10224	2840	2816	4524	4537
5001	5031	1561	1564	3411	3436
388	<800	6877	6790	948	911
832	842	16065	15540	272	<800
165	<800	4076	3837	4359	4227
116	<800	15546	15540	210	<800
4232	4241	1520	1564	4666	4684
6490	6601	141	<800	854	911
3319	3327	83	<800	10084	10002
1081	1076	7283	7183	1420	1410
7379	7393	844	861	2040	2051
907	918	3456	3403	2972	3054
5434	5406	412	<800	4263	4227
14267	14476	4550	4513	4172	4227
7852	7869	439	<800	386	<800
2483	2526	12228	11492	3588	3653
1573	1533	525	<800	267	<800
2706	2706	2095	2044	10609	10500
5788	5722	533	<800	11893	11650
1413	1381	386	<800	645	<800
12127	11968	308	<800	1350	1328
901	918	2561	2598	1765	1742
2140	2117	3867	3837	1002	1007
430	<800	11514	11492	1619	1590
254	<800	743	731	4250	4227
		1621	1564	15992	16197
		720	731	5279	5286
		2767	2816	3965	4227
		144	<800	4595	4684
		542	<800	5928	5910
				900	911
				6373	6450
				1882	1867
				539	<800
				559	<800
				3763	3959

source	1. .166485	/organism="Homo sapiens"	/mol_type="genomic DNA"
Query Match	67.3%;	Score 2605.2;	DB 9; Length 166485;
Best Local Similarity	99.6%;	Pred. No. 0;	
Matches 2634;	Conservative	0; Mismatches 3;	Indels 7; Gaps 2;
QY	1213	AGGAGAAAGTTGTGGAATATGTGGATGACCTCTTTGGAACCTGGAGGAGACCAAGCTAGGCCA	1272
DB	17584	AGGAGAAAGTTGTGGAATATGTGGATGACCTCTTGGAACTGGAGGAGACCAAGCTAGGCCA	17643
QY	1273	CAGCAACCAAGAGACTTCCTCTTGGCGTTTCAGGAAACACAGATTTCTTTGCTTTTCCC	1332
DB	17644	CAGCAACCAAGAGACTTCCTCTTGGCGTTTCAGGAAACACAGATTTCTTTGCTTTTCCC	17703
QY	1333	AACAGCCAGGCTGTGTATACCTCAGAGCCTTCTCTTTACTCTCCAAAGTGAAGGGAAG	1392
DB	17704	AACAGCCAGGCTGTGTATACCTCAGAGCCTTCTCTTTACTCTCCAAAGTGAAGGGAAG	17763
QY	1393	CCCCGCTCTCTAACTGATGATCATCAGGGGTGAGCTGCTTTCCTATCTTCACACCT	1452
DB	17764	CCCCGCTCTCTCTAACTGATGATCATCAGGGGTGAGCTGCTTTCCTATCTTCACACCT	17823
QY	1453	GCCACCTCATGTTTACACCTATCTTTCTCACCTTTTTTTTGGAGATGGAGTCTCGCTCTCT	1512
DB	17824	GCCACCTCATGTTTACACCTATCTTTCTCACCTTTTTTTTGGAGATGGAGTCTCGCTCTCT	17883
QY	1513	TGCCAGGCTGGATGCAATGCAAGTCTCTCAGCTCAGTGCMACTCGGCTCTTGGGTT	1572
DB	17884	TGCCAGGCTGGATGCAATGCAAGTCTCTCAGCTCAGTGCMACTCGGCTCTTGGGTT	17943
QY	1573	CAAGCAATTTCTGTCATCAGCTCCCGAGTACCTGGGATACAGGATGTCGCAACACG	1632
DB	17944	CAAGCAATTTCTGTCATCAGCTCCCGAGTACCTGGGATACAGGATGTCGCAACACG	18003
QY	1633	CCGGCTAATTTTGTATTTTATAGTAGACGGGGTTTTGCCATGTTGGCCAGGCTGGTCT	1692
DB	18004	CCGGCTAATTTTGTATTTTATAGTAGACGGGGTTTTGCCATGTTGGCCAGGCTGGTCT	18063
QY	1693	CGAACTCTTGACATTACAGATGATCCATCTGCTTGGCTCCACAGTGTGGGATTCAGG	1752
DB	18064	CGAACTCTTGACATTACAGATGATCCATCTGCTTGGCTCCACAGTGTGGGATTCAGG	18123
QY	1753	CGTAGGCCACATGCGCGCTCTTCTCACTTACACCTGTCTTCTTATCTCTCAATC	1812
DB	18124	CGTAGGCCACATGCGCGCTCTTCTCACTTACACCTGTCTTCTTATCTCTCAATC	18183
QY	1813	TGTTTTACACCTTCACTCCCTCTCTCTCATGTTTCACTGTCTTCCCTCCCTCCATTA	1872
DB	18184	TGTTTTACACCTTCACTCCCTCTCTCTCATGTTTCACTGTCTTCCCTCCCTCCATTA	18243
QY	1873	GCTGCCCTTTTACCAATTTTGGTTTGAAGGCGAGTCTTCTCGGCTTGTTTTTTGTGTTT	1932
DB	18244	GCTGCCCTTTTACCAATTTTGGTTTGAAGGCGAGTCTTCTCGGCTTGTTTTTTGTGTTT	19303
QY	1933	TCCCGAAATCAGTATATTTTTTAATAGAAACATTCCTAGAGATGATATGT	1992
DB	19304	TCCCGAAATCAGTATATTTTTTAATAGAAACATTCCTAGAGATGATATATGT	18363
QY	1993	GAAGACCTCTTGGCTTATTTGCTTTTCCAGATTTTGTAGTCTCTTCCCATCCGG	2052
DB	18364	GAAGACCTCTTGGCTTATTTGCTTTTCCAGATTTTGTAGTCTCTTCTCCCATCCGG	18423
QY	2053	AAAGATGGTGGAAAGACATAGGCTAAATTTCTCAGCCTCAAGTGGTCTTCACTTGGTCT	2112
DB	18424	AAAGATGGTGGAAAGACATAGGCTAAATTTCTCAGCCTCAAGTGGTCTTCACTTGGTCT	18483
QY	2113	GACTTGTACCAATTTCTAGCACCACTGAAACCAAGTTTGTAGTAGAGTGTAGAGTGCAG	2172
DB	18484	GACTTGTACCAATTTCTAGCACCACTGAAACCAAGTTTGTAGTAGAGTGTAGAGTGCAG	18543
QY	2173	AAATGTGGCTTTTGGCCACTTTGTGATCTTCCAAATATACACGGGTTCGCCGATCCCATTT	2232

```
Db 19624 TTGGTTTATTGTTGGGGTGCTCTCAGAAAGAACCTGTGAAGAGCCTTTATCAGTCATTTA 19583
QY 3306 TTGGCTGTGAGAGTGTCTCTGGAGTGTGGGTACATTTGAAGCGCAAGTCACTTCAGTTGA 3365
Db 19684 TTGGCTGTGAGAGTGTCTCTGGAGTGTGGGTACATTTGAAGCGCAAGTCACTTCAGTTGA 19743
QY 3366 GGGCAAGTCTCTGGAAGAGAGCTGTAGGCATCTGGCAGCTACCATGCGTGTAGTGTGT 3425
Db 19744 GGGCAAGTCTCTGGAAGAGAGCTGTAGGCATCTGGCAGCTACCATGCGTGTAGTGTGT 19803
QY 3426 TGGGGTGGGGTCTCTGGGCACTGTGTGTGAAGGAGTCTGGCAGGCGCACACAGCGCC 3485
Db 19804 TGGGGTGGGGTCTCTGGGCACTGTGTGTGAAGGAGTCTGGCAGGCGCACACAGCGCC 19863
QY 3486 CCTACTGAACCATCAGCATGTGAGTGGCATTTAAAGCATGCGAGCTGGAGGGGCACTG 3545
Db 19864 CCTACTGAACCATCAGCATGTGAGTGGCATTTAAAGCATGCGAGCTGGAGGGGCACTG 19923
QY 3546 AGATTGTCTCTGAGTATTACTGAGAAGCAACAGAAAGAGCCATGGATGGAGCCCTTGGG 3605
Db 19924 AGATTGTCTCTGAGTATTACTGAGAAGCAACAGAAAGAGCCATGGATGGAGCCCTTGGG 19983
QY 3606 CTCCTGGGAATGGGAATACGCAAGGAGTGAAGAGGATTTACCTTAAGGTGAGAGA 3665
Db 19984 CTCCTGGGAATGGGAATACGCAAGGAGTGAAGAGGATTTACCTTAAGGTGAGAGA 20043
QY 3666 AAACCAAGAGAGTGTGGTGTCTGGAAGCTGAGCTTCTTTTATTCACCTCATTCCTTC 3725
Db 20044 AAACCAAGAGAGTGTGGTGTCTGGAAGCTGAGCTTCTTTTATTCACCTCATTCCTTC 20103
QY 3726 TCCAAATAGCCACTGTGTGTAGTGTGGGCCCTCCAGGTTGAAGGCAAGAGGAAAGGC 3785
Db 20104 TCCAAATAGCCACTGTGTGTAGTGTGGGCCCTCCAGGTTGAAGGCAAGAGGAAAGGC 20163
QY 3786 ACAGGCTTTGGGAACAGAGCTTCTGCAATAGCTGGAAGGATTAAGGATAGAG 3845
Db 20164 ACAGGCTTTGGGAACAGAGCTTCTGCAATAGCTGGAAGGATTAAGGATAGAG 20223
QY 3846 TGTT 3849
Db 20224 TGTT 20227

RESULT 4
HSAJ6470
LOCUS Homo sapiens mRNA for cartilage-associated protein (CASP).
ACCESSION AJ006470
VERSION AJ006470.1 GI:3687321
KEYWORDS cartilage-associated protein; CASP.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Tonachini,L., Morello,R., Monticone,M., Skaug,J., Scherer,S.W.,
Cancedda,R. and Castagnola,P.
TITLE cDNA cloning, characterization and chromosome mapping of the gene
encoding human cartilage associated protein (CRTAP)
JOURNAL Cytogenet. Cell Genet. 87 (3-4), 191-194 (1999)
MEDLINE 20169181
PUBMED 10702654
REFERENCE 2 (bases 1 to 2307)
AUTHORS Castagnola,P.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Castagnola P., Biotechnologie in Oncologia,
16132, ITALY
FEATURES
source
1..2307
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
```

```
/clone="Contig of pCPH2 and I.M.A.G.E. cloneID 1090695"
/tissue_type="fetal brain"
/dev_stage="19-23 weeks-old"
1..2307
/gene="casp"
CDS
12..1217
/gene="casp"
/codon_start=1
/product="cartilage-associated protein (CASP)"
/protein_id="CAA07054.1"
/db_xref="GI:3687322"
/db_xref="SPTREMBL:O75718"
/translation="MEPRGRGAALLALICVCAURAGRAQYERYSFRFPDELMPLE
ESYRHALDKYSGEHWAESVGYLEISLRHLRLRSEAFCHRNCSAPOPEAAGLAS
YPLRFLGELLRAHCKKQGLPAFRQSPRSREVIADFORREPKYFLQFAVFRANN
LPKIAAHTFLAKHPDDMMKNMAYKSLPAEDYIKDLETYSVESLFIARVAYN
GENRSTIDMELALPDFKAFVECLAACGSEIREIKDFYLSADHVFVELECKIQ
CBENLTPVIGGVPEKVFATMYHQLQFAYVKLNDLKNAAPCAVSYLLDPNDKVMQON
LVYQVHRDTWGLSDBHFQPRPEAVQFNVTTLQELDYDPAKENIMDDDEGEVEYVD
DLLELETS"
12..89
/gene="casp"
90..1214
/gene="casp"
/product="cartilage-associated protein (CASP)"
BASE COUNT 545 a 633 c 548 g 578 t 3 Others
ORIGIN

Query Match 57.8%; Score 2237; DB 9; Length 2307;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2263; Conservative 3; Mismatches 10; Indels 3; Gaps 2;

QY 53 GCCGGCGCGATGGAGCCGGGGCGCGGGGGCGCGGGGGCGCTGCTAGCCCTGCTGTGCGT 112
Db 2 GCACGAGGCGATGAGCCGGGGCGCGGGGGCGCGGGGGCGCTGCTAGCCCTGCTGTGCGT 61
QY 113 GGCTCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 172
Db 62 GGCTCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 121
QY 173 ACGGGACGAGCTGATGCTCGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 232
Db 122 ACGGGACGAGCTGATGCTCGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 181
QY 233 CGAGCACTCGGCGGAGAGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 292
Db 182 CGAGCACTCGGCGGAGAGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 241
QY 293 GCGCGACGAGGAGCTTCTGCGCACCGCAACTGCGAGCGCGCGCGCGCGCGCGCGCGCG 352
Db 242 GCGCGACGAGGAGCTTCTGCGCACCGCAACTGCGAGCGCGCGCGCGCGCGCGCGCGCG 301
QY 353 CGCGGCGCTCGCAGCTATCCGAGCTGCGCTCTTCCGGGGCGCTGCTGCGCGCGCGCGCA 412
Db 302 CGCGGCGCTCGCAGCTATCCGAGCTGCGCTCTTCCGGGGCGCTGCTGCGCGCGCGCGCA 361
QY 413 CTGCTCAAGCGCTGCAAGCAGGCGCTGCCAGCCCTTCCGCGAGTCCCGAGCCCGAGCGCG 472
Db 362 CTGCTCAAGCGCTGCAAGCAGGCGCTGCCAGCCCTTCCGCGAGTCCCGAGCCCGAGCGCG 421
QY 473 GGTCTGCGGAGCTTCCAGCGCGCGGAGCGCTCAAGTTCCTGCGAGTTCGCTTACTTCAA 532
Db 422 GGTCTGCGGAGCTTCCAGCGCGCGGAGCGCTCAAGTTCCTGCGAGTTCGCTTACTTCAA 481
QY 533 GCGCAATATCTCCCAAGCGCATCGCGCTGCTCACACCTTCTTACTGAGAGCATCTCTGA 592
Db 482 GCGCAATATCTCCCAAGCGCATCGCGCTGCTCACACCTTCTTACTGAGAGCATCTCTGA 541
QY 593 TGACGAATGATGAAGAGGAACTGGCATATTATAAGAGCCTGCTGTGCTGCGGAGGACTA 652
Db 542 TGACGAATGATGAAGAGGAACTGGCATATTATAAGAGCCTGCTGTGCTGCGGAGGACTA 601
QY 653 CATTAAGACCTGGNAACCAAGTCAATATGAAGCCTGTTCAATCGAGCAGTCCGGGCGATA 712
```

[illegible]

Db	1682	TCCACAGTGTGGGATTACAGCGGTGAGCCACCATGCCCGGCGTCTTTCTCACCTTTAC	1741
Qy	1790	ACCTGTCTTCTATPCCFCACATGTTTTCACACCCCTCATCCCTGTCTTCTCATGTTC	1849
Db	1742	ACCTGTCTTCTATPCCFCACATGTTTTCACACCCCTCATCCCTGTCTTCTCATGTTC	1801
Qy	1850	CATTGTCTTCCCATGTTTCATAGTGCCTTTCTTACCATTTTGGTTTGAAGGCGAGTCT	1909
Db	1802	CATTGTCTTCCCATGTTTCATAGTGCCTTTCTTACCATTTTGGTTTGAAGGCGAGTCT	1861
Qy	1910	TCTCTGCTTGTCTTTTGTCTTTTCCAGAAAAATCAGTATTATTTTAAATTAAGAAAA	1969
Db	1862	TCTCTGCTTGTCTTTTGTCTTTTCCAGAAAAATCAGTATTATTTTAAATTAAGAAAA	1921
Qy	1970	CATTCTAGAGATGATAATGTGAAAAACCTCTTTGGCTTATTTGCTTTCCAGATTTT	2029
Db	1922	CATTCTAGAGATGATAATGTGAAAAACCTCTTTGGCTTATTTGCTTTCCAGATTTT	1981
Qy	2030	AGTCTCTTCTTCCCATCCGGAAGATGCTGGAGACATAGCTAAATTTCTCCAGCC	2089
Db	1982	AGTCTCTTCTTCCCATCCGGAAGATGCTGGAGACATAGCTAAATTTCTCCAGCC	2041
Qy	2090	TCACATGGTCTTCACTTGTCTGACTTGTACCAATTTAGACCCACTGAAAAACAAGT	2149
Db	2042	TCACATGGTCTTCACTTGTCTGACTTGTACCAATTTAGACCCACTGAAAAACAAGT	2101
Qy	2150	TGAGTAGAGAGTGTAGAGTCAGAAATGTGGCTTTTGGCCCACTTTGCATCTCCAAAAT	2209
Db	2102	TGAGTAGAGAGTGTAGAGTCAGAAATGTGGCTTTTGGCCCACTTTGCATCTCCAAAAT	2161
Qy	2210	ACAACGGTTGGCCGATCCCACTTTGAGGACAAATGCTTAGTTATAAGTCTCCGAGTTGGAAA	2269
Db	2162	ACAACGGTTGGCCGATCCCACTTTGAGGACAAATGCTTAGTTATAAGTCTCCGAGTTGGAAA	2221
Qy	2270	AGGAAGAAAGCCAGAGCTGTCTAGTTTCATTCATCTTCTTCAGTAAATATTTATGAGTA	2328
Db	2222	AGGAAGAAAGCCAGAGCTGTCTAGTTTCATTCATCTTCTTCAGTAAATATTTATGAGTA	2280
RESULT 5			
AK075401		2010 bp	mrna
LOCUS	AK075401	Homo sapiens cDNA PSEC0091 fis, clone PLACE1001683, highly similar to Homo sapiens mRNA for cartilage-associated protein.	PRI 03-SEP-2002
DEFINITION	AK075401		
ACCESSION	AK075401.1	GI:227616465	
VERSION			
KEYWORDS		Cligo capping; fis (full insert sequence).	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		Oka, T., Nishikawa, T., Suzuki, Y., Kawai-Hio, Y., Hayashi, K., Ishii, S., Saito, K., Yamamoto, J., Wakamatsu, A., Nagai, T., Nakamura, Y., Sato, H., Negahari, K., Sugano, S. and Isogai, T.	
TITLE		HRI human cDNA sequencing project	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 2010)	
AUTHORS		Isogai, T. and Yamamoto, J.	
TITLE		Direct Submission	
JOURNAL		Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1536-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)	
COMMENT		HRI human cDNA sequencing project; cDNA 5'- & 3'-end one pass sequencing, clone selection and full insert sequencing; Helix Research Institute (supported by Japan Key technology Center etc.); cDNA library construction; Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center.	
FEATURES		Location/Qualifiers	
source		1. .2010	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	

[illegible]











204 QY TACCGCACCGCTGGACAAAGTACAGCGGCGAGCACTGGGCCGAGAGCSTKGGCTACCTG 263  
181 Db TACCGCACCGCTGGACAAAGTACAGCGGCGAGCACTGGGCCGAGAGC----- 229  
264 QY GAGATCAGCTGGCGCTGCACCGCTTGGTGGGCGAGAGAGCGCTTCTGCCACCGGAAC 323  
230 Db -----TGGCGACAGCGAGGCGCTTCTGCCACCGGAAC 261  
324 QY TGCAGCGCCGCGCGCAGCCGAGCCGCGCGCGGCTCGCCAGCTATCCGAGCTGCGC 383  
262 Db TGCAGCGCGCGCGCAGCCGAGCCGCGCGCGGCTCGCCAGCTATCCGAGCTGCGC 321  
384 QY CTCTTGGGCGCTGCTGCGCGCGCAGCTGCGCTCAAGCGCTGCAAGCAGGCGCTGCCA 443  
322 Db CTCTTGGGCGCGCTGCTGCGCGCGCAGCTGCGCTCAAGCGCTGCAAGCAGGCGCTGCCA 381  
444 QY GCCTTCCGCGCAGTCCCAAGCCAGCCGCGAGGTGCTGGCGACTTCCAGCGCGCGAGGCC 503  
382 Db GCCTTCCGCGCAGTCCCAAGCCAGCCGCGAGGTGCTGGCGACTTCCAGCGCGCGAGGCC 441  
504 QY TACAAGTTCCTGCGAGTTCGCTTACTTCAAGGCAATAATCTCCCAAAGCCATCGCGCT 563  
442 Db TACAAGTTCCTGCGAGTTCGCTTACTTCAAGGCAATAATCTCCCAAAGCCATCGCGCT 501  
564 QY GCTCAGACCTTTCTACTGAAGCATCTGATGAGAAATGATGAAGAGGAACATGGCATAT 623  
502 Db GCTCAGACCTTTCTACTGAAGCATCTGATGAGAAATGATGAAGAGGAACATGGCATAT 561  
624 QY TATAAGAGCCTGCTGGTGGCGGAGGACTACATTAAGAGCCTGGAAACCAAGTCATATGAA 683  
562 Db TATAAGAGCCTGCTGGTGGCGGAGGACTACATTAAGAGCCTGGAAACCAAGTCATATGAA 621  
684 QY AGCTGTTCATCCGAGCAGTGGCGGATACAACGGTGAAGAACTGGAGAAATCCATCACA 743  
622 Db AGCTGTTCATCCGAGCAGTGGCGGATACAACGGTGAAGAACTGGAGAAATCCATCACA 681  
744 QY GACATGAGCTGCGCTTCCGAGCTTCTCAAGAGCCTTTACAGTGTCTCCGAGCGCTGC 803  
682 Db GACATGAGCTGCGCTTCCGAGCTTCTCAAGAGCCTTTACAGTGTCTCCGAGCGCTGC 741  
804 QY GAGGGTTCAGGAGAGTCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAGATCATAT 863  
742 Db GAGGGTTCAGGAGAGTCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAGATCATAT 801  
864 QY GTAGAAGTTCGAATGCAGAAATACAGTGTGAAGAGAACTCACCAGGATATAGAGGC 923  
802 Db GTAGAAGTTCGAATGCAGAAATACAGTGTGAAGAGAACTCACCAGGATATAGAGGC 861  
924 QY TATCCGGTTGAGAAATTTGGGTACCATGTATCATTTACTTGCAGTGTGGCTATTATAAG 983  
862 Db TATCCGGTTGAGAAATTTGGGTACCATGTATCATTTACTTGCAGTGTGGCTATTATAAG 921  
984 QY TTGAACGACCTGAAGAAATGAGCCCTGTGCGAGTCAGCTATCTGCTTTGATCAGAAAT 1043  
922 Db TTGAACGACCTGAAGAAATGAGCCCTGTGCGAGTCAGCTATCTGCTTTGATCAGAAAT 981  
1044 QY GACAAGTTCATGAGCAGAAACCTGGTGTATTACAGTACACAGGAGCACTTGGGGCCTC 1103  
982 Db GACAAGTTCATGAGCAGAAACCTGGTGTATTACAGTACACAGGAGCACTTGGGGCCTC 1041  
1104 QY TCGGATGAGCACTTCCAGCCGAGACCTTGAAGCAGTTCAGTTCCTTTAATGTGACCACTC 1163  
1042 Db TCGATGAGCACTTCCAGCCGAGACCTTGAAGCAGTTCAGTTCCTTTAATGTGACCACTC 1101  
1164 QY CAGAAGGAGCTGATGACTTTGTTAGGAAATATATGATGATGATGAGGAGAGT 1223  
1102 Db CAGAAGGAGCTGATGACTTTGTTAGGAAATATATGATGATGATGAGGAGAGT 1161  
1224 QY GTGGAATATGTGATGACCTTCTGGAACTGGAGAGCAGCAGCTAGCCACAGCAACAAA 1283  
1162 Db GTGGAATATGTGATGACCTTCTGGAACTGGAGAGCAGCAGCTAGCCACAGCAACAAA 1221

1284 QY GAGACTTCTCTTGGCGTTTCAGGAAACACAGATTCTTTGTCTTTTCCCAACAGCCAGG 1343  
1222 Db GAGACTTCTCTTGGCGTTTCAGGAAACACAGATTCTTTGTCTTTTCCCAACAGCCAGG 1281  
1344 QY CTGTTGATACCTCAGAGCCTTCTCTTTACTCTCCAAAGTGAAGGAAAGCCCTCTCTCT 1403  
1282 Db CTGTTGATACCTCAGAGCCTTCTCTTTACTCTCCAAAGTGAAGGAAAGCCCTCTCTCT 1341  
1404 QY CTAACTGATGTATCATCAGGGGTGAGCCTGCTTCTATCTTTCACACTGCACTGCACTCATG 1463  
1342 Db CTAACTGATGTATCATCAGGGGTGAGCCTGCTTCTTCTATCTTTCACACTGCACTCATG 1401  
1464 QY TTCACACCTATCTTTCTCACCTTTTGTGAGTGGAGTCTCGCTCTCTTCCCGCAGGCTG 1523  
1402 Db TTCACACCTATCTTTCTCACCTTTTGTGAGTGGAGTCTCGCTCTCTTCCCGCAGGCTG 1461  
1524 QY GAGTGCATGACAGCTTCTCAGCTCAGTCACTGCAACCTCCGCTCTTGGGTTCAGCAATTCT 1583  
1462 Db GAGTGCATGACAGCTTCTCAGCTCAGTCACTGCAACCTCCGCTCTTGGGTTCAGCAATTCT 1521  
1584 QY GCTGCATCAGCCTCCCGAGTACCTGGGATTAACAGGCACTGTGCCACACCGCCGGCTAATT 1643  
1522 Db GCTGCATCAGCCTCCCGAGTACCTGGGATTAACAGGCACTGTGCCACACCGCCGGCTAATT 1581  
1644 QY TTGATATTTTAGTAGAGAGCGGGTTTGGCAATGTTGGCCAGCTGGTCTCGAACTCTTGA 1703  
1582 Db TTGATATTTTAGTAGAGAGCGGGTTTGGCAATGTTGGCCAGCTGGTCTCGAACTCTTGA 1641  
1704 QY CTTTCAGATGATCCATCTGCTTGGCTCCCACTGCTGGGATTAACAGGCGTGAGCCAC 1763  
1642 Db CTTTCAGATGATCCATCTGCTTGGCTCCCACTGCTGGGATTAACAGGCGTGAGCCAC 1701  
1764 QY ATGCCCGGCTCTTCTCACCTTTACACCTGTCTTCTTATCTCCTCACATCTGTTTTCAC 1823  
1702 Db ATGCCCGGCTCTTCTCACCTTTACACCTGTCTTCTTATCTCCTCACATCTGTTTTCAC 1761  
1824 QY CTTTCATCTCTGCTCTCTCTCATGTTTCACATGTTCTTCCCATGTTTCATAGTCTTCT 1883  
1762 Db CTTTCATCTCTGCTCTCTCTCATGTTTCACATGTTCTTCCCATGTTTCATAGTCTTCT 1821  
1884 QY TACCATTTTGGTTTGAAGGCGAGTCTTCTCTGGCTGTTGTTTTTGTTCACAGAAAT 1943  
1822 Db TACCATTTTGGTTTGAAGGCGAGTCTTCTCTGGCTGTTGTTTTTGTTCACAGAAAT 1881  
1944 QY CAGTATATTTTAAATAGAAAACATTCCTTAGAAGATG 1984  
1882 Db CAGTATATTTTAAATAGAAAACATTCCTTAGAAGATG 1922

RESULT 9  
AK092990  
LOCUS  
DEFINITION Homo sapiens cDNA FLJ35671 fis, clone SPLEN2018180, highly similar to Homo sapiens mRNA for cartilage-associated protein.  
ACCESSION AK092990  
VERSION AK092990.1 GI:21751708  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S., Fukuyama,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yanashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Suzano,S., Nagahari,K., Masubo,Y., Nagai,K. and Isogai,T.  
TITLE NED0 human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1903)

AUTHORS	Isogai, T. and Yamamoto, J.	
TITLE	Direct Submission	
JOURNAL	Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7	
COMMENT	Kazusa-Kamatari, Kibazaru, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.	
FEATURES	Location/Qualifiers	
source	1..1903	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/clone="SPLEN2018180"	
	/tissue_type="spleen"	
	/clone_lib="SPLEN2"	
	/note="cloning vector: pME18SFL3"	
BASE COUNT	421 a 553 c 460 g 459 t	
ORIGIN		
Query Match	46.5%; Score 1798.8; DB 9; Length 1903;	
Best Local Similarity	96.6%; Pred. No. 1.9e-301;	
Matches 1888; Conservative	2; Mismatches 9; Indels 55; Gaps 3;	
QY	47 CTTTTCGGCGGCGGATGGAGCGGGCGCGGGGGCGCGCGTGTAGCGCTGCT	106
DB	1 CTTTTCGGCGGCGGATGGAGCGGGCGCGGGGGCGCGCGTGTAGCGCTGCT	59
QY	107 GTGCGTGCCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	166
DB	60 GTGCGTGCCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	119
QY	167 CTTCCACGGGACGAGCTGATGCGCGCTCGAGTGCCTACCGGACGCGCTGGACAGTA	226
DB	120 CTTCCACGGGACGAGCTGATGCGCGCT-----	146
QY	227 CAGCGCGGACACTGGCGCGGAGCGSTXGGGTACTCTGGAGATCAGCTTGGCGGCTGCACCG	286
DB	147 -----CGAGAGCTGGGTACTCTCTGAGATCAGCTTGGCGGCTGCACCG	188
QY	287 CTTGTGCGCGACGACGAGCGCTTTCGCGACCGGAACTGAGCGCGCGCGCGCGCGCGA	346
DB	189 CTTGTGCGCGACGACGAGCGCTTTCGCGACCGGAACTGAGCGCGCGCGCGCGCGA	248
QY	347 GCCCGCGCGCGCTCGCGAGCTATCCGAGTGGCGCTCTTGGGGGCGCTGTGCGCGG	406
DB	249 GCCCGCGCGCGCTCGCGAGCTATCCGAGTGGCGCTCTTGGGGGCGCTGTGCGCGG	308
QY	407 CGCGCACTGCCTCAAGCGCTGCAAGCGGCGCTTGCAGCGCTTGCAGCGCTTGCAGCGCG	466
DB	309 CGCGCACTGCCTCAAGCGCTGCAAGCGGCGCTTGCAGCGCTTGCAGCGCGCTTGCAGCGCG	368
QY	467 CGCGAGGTGCTGGCGGACTTCCAGCGCGCGAGCGCTTACAGTTCCTGCAAGTTCGCTTA	526
DB	369 CGCGAGGTGCTGGCGGACTTCCAGCGCGCGAGCGCTTACAGTTCCTGCAAGTTCGCTTA	428
QY	527 CTTCAAGGCAATAATCTCCCAAGCCATCGCGCTGCTCACACCTTCTTACTGAAGCA	586
DB	429 CTTCAAGGCAATAATCTCCCAAGCCATCGCGCTGCTCACACCTTCTTACTGAAGCA	488
QY	587 TCCTGATGACGAATGATGAGAGGAAATGCGCATATTTAAGAGCTTGCCTGCGCGA	646
DB	489 TCCTGATGACGAATGATGAGAGGAAATGCGCATATTTAAGAGCTTGCCTGCGCGA	548
QY	647 GGACTACATTAAGACCTTGGAAACCAAGTATATGAAAGCGCTTTCATCCGAGCAGTGG	706
DB	549 GGACTACATTAAGACCTTGGAAACCAAGTATATGAAAGCGCTTTCATCCGAGCAGTGG	608
QY	707 GGCATACAAACGCTGAGAACTGGAGAACTCCATACAGACATGAGCTGGCCCTTCCCGA	766
DB	609 GGCATACAAACGCTGAGAACTGGAGAACTCCATACAGACATGAGCTGGCCCTTCCCGA	668
QY	767 CTTCTTCAAAGCGCTTTTACGAGTGTCTCGACGCTTCCAGGGTTCAGGGAGATCAAGA	826
DB	669 CTTCTTCAAAGCGCTTTTACGAGTGTCTCGACGCTTCCAGGGTTCAGGGAGATCAAGA	728
QY	827 CTTCAAGGATTTCTACCTTCCATAGCAGATCATTTATGTAGAACTTCTGGAATGCAAAAT	886
DB	729 CTTCAAGGATTTCTACCTTCCATAGCAGATCATTTATGTAGAACTTCTGGAATGCAAAAT	788
QY	887 ACAGTGTGAAGAAACCTCACCCAGTTATAGAGGCTATCCGGTTGAGAAATTTGTGGC	946
DB	789 ACAGTGTGAAGAAACCTCACCCAGTTATAGAGGCTATCCGGTTGAGAAATTTGTGGC	848
QY	947 TACCATGATATCACTTCTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTG	1006
DB	849 TACCATGATATCACTTCTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTG	908
QY	1007 CCGCTGTGAGTCACTATCTCTTTTATGATCAGAAATGACAAAGTGTCTGAGCAAACT	1066
DB	909 CCGCTGTGAGTCACTATCTCTTTTATGATCAGAAATGACAAAGTGTCTGAGCAAACT	968
QY	1067 GGTGTATTACAGTACACAGGAGCACTTGGGGCTCTCGGATGAGCACTTCCAGCCAG	1126
DB	969 GGTGTATTACAGTACACAGGAGCACTTGGGGCTCTCGGATGAGCACTTCCAGCCAG	1028
QY	1127 ACCTGAAGCAGTTTCTTTTAAATGTGACCACTCCAGAGGAGCTGTATGACTTTGC	1186
DB	1029 ACCTGAAGCAGTTTCTTTTAAATGTGACCACTCCAGAGGAGCTGTATGACTTTGC	1088
QY	1187 TAAGGAAATATATGAGTGTGAGGAGAAAGTGTGGAATATGGAATGAGTGTCTT	1246
DB	1089 TAAGGAAATATATGAGTGTGAGGAGAAAGTGTGGAATATGGAATGAGTGTCTT	1148
QY	1247 GGAACCTGGAGGAGACAGCTAGCCACCAACCAAGAGACTTCTCTTGGCGTTTCAGG	1306
DB	1149 GGAACCTGGAGGAGACAGCTAGCCACCAACCAAGAGACTTCTCTTGGCGTTTCAGG	1208
QY	1307 AAACACAGATTCTTTTGTCTTTTCCAAAGCGGCTGTGATACCTCAGAGCTTCT	1366
DB	1209 AAACACAGATTCTTTTGTCTTTTCCAAAGCGGCTGTGATACCTCAGAGCTTCT	1268
QY	1367 CTTTACTCTCCAAAGTGAAGGAGAGCGCCGCTCTCTTAAGTGTATGATGATGATGATG	1426
DB	1269 CTTTACTCTCCAAAGTGAAGGAGAGCGCCGCTCTCTTAAGTGTATGATGATGATGATG	1328
QY	1427 AGCTTGCCTTTCTTATCTTACACCTGCGCACTTGTTCACACCTATCTTTTCTCAC	1484
DB	1329 AGCTTGCCTTTCTTATCTTACACCTGCGCACTTGTTCACACCTATCTTTTCTCAC	1388
QY	1485 -TTTTTTTTGAGATGAGTCTGCTCTCTTGGCGAGGCTGAGTGAATGGACGTTCTC	1543
DB	1389 TTTTTTTTTGAGATGAGTCTGCTCTCTTGGCGAGGCTGAGTGAATGGACGTTCTC	1448
QY	1544 AGCTCACTGCAACCTCCGCTCTTGGGTTCAAGCAATTTCTGCTGATCAGCTCCCGAGT	1603
DB	1449 AGCTCACTGCAACCTCCGCTCTTGGGTTCAAGCAATTTCTGCTGATCAGCTCCCGAGT	1508
QY	1604 ACCTGGGATACAGGATGTGCGCAACGCGCGGCTTAATTTGTATTTTAGTAGAGAG	1663
DB	1509 ACCTGGGATACAGGATGTGCGCAACGCGCGGCTTAATTTGTATTTTAGTAGAGAG	1568
QY	1664 GGGTTTGGCATGTTGGCGAGGCTGTGCTCGAACTCTTGACTTCAGATGATCCATCTGCC	1723
DB	1569 GGGTTTGGCATGTTGGCGAGGCTGTGCTCGAACTCTTGACTTCAGATGATCCATCTGCC	1628
QY	1724 TTGGCTTCCACAGTGTGGGATTAAGCGGTGAGCCACCAATGCGCGGCTTTTCTCTAC	1783
DB	1629 TTGGCTTCCACAGTGTGGGATTAAGCGGTGAGCCACCAATGCGCGGCTTTTCTCTAC	1688
QY	1784 CTTTACACCTGTCTTCTTATCTCTACATCTGTTTTCACACCTTTCATCCCTGCTTCTCA	1843

```

1689 CTTTACACCTGCTCTTCTTATCTCCTCACATCTGTTTTCACAGCTTCATCCCTGCTCTCTCA 1748
1844 TGTTACACACTTGTCTTCCCATGTTCTAGTGTGCTTCTTACCATTTTGGTTGAAGGG 1903
1749 TGTTACACACTTGTCTTCCCATGTTCTAGTGTGCTTCTTACCATTTTGGTTGAAGGG 1808
1804 CAGTCTTCTCTGGCTGTGTTTTTTTGTGTTTTTCCAGAAAATCAGTATTATTTTAAATAA 1963
1809 CAGTCTTCTCTGGCTGTGTTTTTTGTTTTTCCAGAAAATCAGTATTATTTTAAATAA 1868
1964 GAAAACATCTCTAGAGATGATATTGTGAAA 1997
1869 GAAAACATCTCTAGAGATGATGAAAAA 1902

RESULT 10
LOCUS AK097650 1810 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ40331 fis, clone TESR12031687, highly similar
to Homo sapiens mRNA for cartilage-associated protein.
ACCESSION AK097650
VERSION AK097650.1 GI:21757488
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Kanehori, K., Ishibashi, T., Chiba, Y., Fujimori, K., Hiraoka, S.,
Tanai, H., Watanabe, S., Ishida, S., Ono, Y., Hotuta, T., Watanabe, M.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
Yamamoto, J., Isono, K., Kawai-Hio, Y., Saito, K., Nishikawa, T.,
Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
Wagatsuma, M., Takahashi-Fujii, A., Oshima, A., Sugiyama, A.,
Kawakami, E., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K.
and Isogai, T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1810)
AUTHORS Isogai, T. and Yamamoto, J.
DIRECT SUBMISSION
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5' - & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
Location/Qualifiers
1..1810
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TESR12031687"
/cisue_type="testis"
/clone_lib="TEST12"
/clone="cloning vector: pME18SFL3"
BASE COUNT 375 a 550 c 449 g 436 t
ORIGIN
Query Match 42.8%; Score 1656; DB 9; Length 1810;
Best Local Similarity 93.1%; Pred. No. 9.9e-277;
Matches 1806; Conservative 2; Mismatches 2; Indels 130; Gaps 2;
39 CTTCTCTTCTTCTCCCGGCGGATGAGCCGGGGCGGGGGCGGGCGGCGCTGCTA 98
1 CTTCTCTTCTTCTCCCGGCGGCGATGAGCCGGGGCGGGGGCGGGCGGCGCTGCTA 60

```

```

99 GGGCTGCTGTGGTGGCTGGCGCTGGCGCGCGGGCGGCCAATACAGACGTACAGC 158
61 GGGCTGCTGTGGTGGCTGGCGCTGGCGCGCGGGCGGCCAATACAGACGTACAGC 120
159 TTCCGACAGTTTCCACAGGACGAGCTGATGCCGCTCGAGTCGGCTACGGGACGCGCTG 218
121 TTCCGACAGTTTCCACAGGACGAGCTGATGCCGCTCGAGTCGGCTACGGGACGCGCTG 180
219 GACAAAGTACAGCGGGGAGCACTGGGCGGAGAGCSTKGGCTACTCTGAGATCAGCCTGGGG 278
181 GACAAAGTACAGCGGGGAGCACTGGGCGGAGAGCTGGGCGGAGAGCTGAGATCAGCCTGGG 240
279 CTGCAACGCTTGTCTGGCGGACAGCGAGGCTTCTCCACCGCAACTGCAGCGCCGCGCG 338
241 CTGCAACGCTTGTCTGGCGGACAGCGAGGCTTCTCCACCGCAACTGCAGCGCCGCGCG 300
339 CAGCCCGAGCCCGCGCGGCTCGCAGCTATCCGAGCTCGGCTCTTCTGGGGGGCGCTG 398
301 CAGCCCGAGCCCGCGCGGCTCGCAGCTATCCGAGCTGCGCTCTTCTGGGGGGCGCTG 360
399 CTGGCGCGCGCAGCTGCTCAAGCGCTGCAAGCAGGCGCTGCCAGCTTCCGCGCAGTCC 458
361 CTGGCGCGCGCAGCTGCTCAAGCGCTGCAAGCAGGCGCTGCCAGCTTCCGCGCAGTCC 420
459 CAGCCCGAGCCCGCGGAGTGTCTGGCGGACTTCCAGCGCGCGGCGCTTACAAGTTCCTGCG 518
421 CAGCCCGAGCCCGGAGTGTCTGGCGGACTTCCAGCGCGCGGCGCTTACAAGTTCCTGCG 480
519 TTGCTTTACTCAAGGCAAAATAATCTCCCAAGGCAATCGCGCTGCTCACACCTTTCTA 578
481 TTGCTTTACTCAAGGCAAAATAATCTCCCAAGGCAATCGCGCTGCTCACACCTTTCTA 540
579 CTGAACATCTGTATGACGAATATGTAAGAGGAATATGATATTAAGAGCTGCTGCT 638
541 CTGAACATCTGTATGACGAATATGTAAGAGGAATATGATATTAAGAGCTGCTGCT 600
639 GGTGCGGAGGACTACATTAAGACCTTGAAGAACCAAGTATGTAAGAGCTGCTTCAATCGGA 698
601 GGTGCGGAGGACTAC-TTAAAGACCTTGAAGAACCAAGTATGTAAGAGCTGCTTCAATCGGA 659
699 GAGTGGCGGACATACAAACGCTGAGAAATCGGAGAAATCCATCACAGACATGAGAGCTGGCC 758
660 GAGTGGCGGACATACAAACGCTGAGAAATCGGAGAAATCCATCACAGACATGAGAGCTGGCC 719
759 CTTCCGAGCTTCTTCAAGGCTTTTACGAGTGTCTCGAGCGCTCGAGGCTTCCAGGAG 818
720 CTTCCGAGCTTCTTCAAGGCTTTTACGAGTGTCTCGAGCGCTCGAGGCTTCCAGGAG 779
819 ATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAGATCATTTATGTAAGAGTCTTGGAA 878
780 ATCAAGGACTTCAAGGATTTCTACCTTTCCATAGCAG----- 816
879 TGCAAAATACAGTGTGAAGAGAACCTCACCCAGTTATAGGAGGCTATCCGCTTGAGAA 938
817 ----- 816
939 TTTTGGCTACCATGTATCATTTACTTTGAGTTTGGCTATTATAAGTTGAACACCTGAAG 998
817 -----TGAACGACCTGAAG 830
999 AATGCAAGCCCTGTGTCAGTACGATATCTGCTTTTGTATCAGAATGACAGTCTATGCG 1058
831 AATGCAAGCCCTGTGTCAGTACGATATCTGCTTTTGTATCAGAATGACAGTCTATGCG 890
1059 CAGAACTCTGTATTATACAGTACCAAGGACACTTGGGGCTCTCGATCAGCACTTC 1118
891 CAGAACTCTGTATTATACAGTACCAAGGACACTTGGGGCTCTCGATGAGCACTTC 950
1119 CAGCCAGACCTGAAGCAGTTCAGTCTTTTAAATGACCACTCCAGAGAGAGTGTAT 1178
951 CAGCCAGACCTGAAGCAGTTCAGTCTTTTAAATGACCACTCCAGAGAGAGTGTAT 1010
1179 GACTTTGCTTAAGAAAAATATAATGATGATGAGGAGAAATTTGTGGAATATGTGGAT 1238

```

Db 1011 GACCTTCTAGGAAATATATGGATGATGAGGAGAACTTGTGGATATGGAT 1070  
Qy 1239 GACCTTCTAGGAGAGACAGCTAGCCACAGCAACAAAGAGACTTCTCTTGG 1298  
Db 1071 GACCTTCTAGGAGAGACAGCTAGCCACAGCAACAAAGAGACTTCTCTTGG 1130  
Qy 1299 CGTTGAGAAACACAGATTTCTTGTCTTTTCCACAGCCAGGCTGTGATACCTCAG 1358  
Db 1131 CGTTGAGAAACACAGATTTCTTGTCTTTTCCACAGCCAGGCTGTGATACCTCAG 1190  
Qy 1359 AGCTTCTCTTACTCTCCAAAGTGAAGGAGAGCCCTCTCTCTTAATGATGATCAT 1418  
Db 1191 AGCTTCTCTTACTCTCCAAAGTGAAGGAGAGCCCTCTCTCTTAATGATGATCAT 1250  
Qy 1419 CAGGGGTGAGCTGCTCTCTTCTTCACTTCACTGCGACCTCATGTTTCAACCTACTTCT 1478  
Db 1251 CAGGGGTGAGCTGCTCTCTTCTTCACTTCACTGCGACCTCATGTTTCAACCTACTTCT 1310  
Qy 1479 CTCACCTTTTTTTTGTAGATGAGTCTGCTCTCTTCCAGGCTGAGTGCATGACG 1538  
Db 1311 CTCACCTTTTTTTTGTAGATGAGTCTGCTCTCTTCCAGGCTGAGTGCATGACG 1370  
Qy 1539 TTCTCAGCTCACTGCACTCCGCTCTTGGTTTCAAGCAATCTCTGATCAGCTCC 1598  
Db 1371 TTCTCAGCTCACTGCACTCCGCTCTTGGTTTCAAGCAATCTCTGATCAGCTCC 1430  
Qy 1599 CGAGTACCTGGATTCAGGCTGTCACACACCCCGGCTAAATTTTGTATTTTAGTAG 1658  
Db 1431 CGAGTACCTGGATTCAGGCTGTCACACACCCCGGCTAAATTTTGTATTTTAGTAG 1490  
Qy 1659 AGACGGGTTTGGATGTTGGCCAGGCTGCTTCAAGCTCTTCACTTCAAGTATCCAT 1718  
Db 1491 AGACGGGTTTGGATGTTGGCCAGGCTGCTTCAAGCTCTTCACTTCAAGTATCCAT 1550  
Qy 1719 CTGCTTGGCTTCCACAGCTGCTGGATTACAGCGCTGAGCCACCATGCCGCTCTTT 1778  
Db 1551 CTGCTTGGCTTCCACAGCTGCTGGATTACAGCGCTGAGCCACCATGCCGCTCTTT 1610  
Qy 1779 CTCACCTTTACCTGCTCTTATTCCTCACTCTGTTTTCACCTTCACTTCACTTCTT 1838  
Db 1611 CTCACCTTTACCTGCTCTTATTCCTCACTCTGTTTTCACCTTCACTTCACTTCTT 1670  
Qy 1839 CCTCATGTCACACTGCTTCCCGCTTTCATAGCTGCTTCTTCACTTCACTTCTTCTG 1898  
Db 1671 CCTCATGTCACACTGCTTCCCGCTTTCATAGCTGCTTCTTCACTTCACTTCTTCTG 1730  
Qy 1899 AAGGCGACTTCTCTGCTTGTGTTTTTTTGTGTTTTTCCAGAAATCAGTATTTT 1958  
Db 1731 AAGGCGACTTCTCTGCTTGTGTTTTTTTGTGTTTTTCCAGAAATCAGTATTTT 1790  
Qy 1959 AATAAGAAAACATTCCTAG 1978  
Db 1791 AATAAGAAAACATTCCTAG 1810

RESULT 11  
BC008745  
LOCUS BC008745 1439 bp mRNA linear PRI 12-JUL-2001  
DEFINITION Homo sapiens, cartilage associated protein, clone MGC:1926  
IMAGE:3347384, mRNA, complete cds.  
ACCESSION BC008745  
VERSION BC008745.1 GI:14250580  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1439)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland,  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nri.nih.gov](mailto:nisc_mgc@nri.nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,  
Tongson, B.E., Touchman, J.W., Teursson, C., Vogt, J.L., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 5 Row: i Column: 24  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 3687321.  
Location/Qualifiers  
1. 1439  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="LocusID:10491"  
/db\_xref="taxon:9606"  
/clone="MGC:1926 IMAGE:3347384"  
/tissue\_type="Kidney, renal cell adenocarcinoma"  
/clone\_lib="NIH MGC 14"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
36. 1241  
/codon\_start=1  
/product="Cartilage associated protein"  
/protein\_id="AAH08745.1"  
/db\_xref="GI:14250581"  
/translation="MEPGRGAALALLCVACLRAGRAQYYSFRSPRDELMP  
ESAYHALDKYSGEHWASVGLTISLRHLRLDSEAFCHRNCSAAPPQFAAGLAS  
YVELFLGGLARHAKLRCXGGLPAPQSOVSRDLADFORPQKLPFLQFAYFRKAN  
LPKAAHAATTELLKHPDDEMKRWAVYKSLPGAEVYTKDLETYSYELFTRAVRAYN  
GENWTSITDMELALPDKAFYECLAAECSESRKPKDFYLSIADHYVLEKIQ  
CEENLTPVIGGYPVEKFAVTHYLQFAYYKLNLDKNAAPCAVSLFLDQNDKVMQON  
LVYYQYHRDTWGLSDEHPQPRPEAVQFENVTTLOKELYFAKENIMDDGEVVEYVD  
DLLELEETS"

BASE COUNT 362 a 415 c 375 g 287 t  
ORIGIN  
Query Match 35.4%; Score 1370.4; DB 9; Length 1439;  
Best Local Similarity 99.3%; Pred. No. 2.8e-227;  
Matches 1374; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
Qy 37 TTCCTTCTCTCTTCGCGCGCGCGATGAGCCGCGCGCGCGCGCGCGCGCTGC 96  
Db 10 TTCCTTCTCTCTTCGCGCGCGCGATGAGCCGCGCGCGCGCGCGCGCTGC 69  
Qy 97 TAGCGCTGTGTGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCTACA 156  
Db 70 TAGCGCTGTGTGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCTACA 129  
Qy 157 GCTTCGCGAGCTTCCACGCGGACGAGCTGATCCGCTCGAGTCGCGCTACCGGACGCGC 216  
Db 130 GCTTCGCGAGCTTCCACGCGGACGAGCTGATCCGCTCGAGTCGCGCTACCGGACGCGC 189  
Qy 217 TGGACAAGTACAGCGCGGAGCACTGGCGCGAGAGACTTGGCTTACTGGAGATCACTGCTGC 276  
Db 190 TGGACAAGTACAGCGCGGAGCACTGGCGCGAGAGCGTGGGCTTACTGGAGATCACTGCTGC 249



```
QY 277 GGCTGACCGCTTGCTGCGCGACAGCGAGCGCTTCTGCCACCGCACTGCAGCGCGCGC 336
Db 250 GGTGACCGCTTGCTGCGCGACAGCGAGCGCTTCTGCCACCGCACTGCAGCGCGCGC 309
QY 337 CGCAGCGCGCGCGCGCGCGCTGCGAGCTATCCGAGCTGCGCTCTTTCCGGGGCC 396
Db 310 CGCAGCGCGCGCGCGCGCGCTGCGAGCTATCCGAGCTGCGCTCTTTCCGGGGCC 369
QY 397 TGTGCGCGCGCGCGCGCTCAAGCGTGAAGCAGCGCGCTGCCAGCTTCCGCCAGT 456
Db 370 TGTGCGCGCGCGCGCGCTCAAGCGTGAAGCAGCGCGCTGCCAGCTTCCGCCAGT 429
QY 457 CCAGCGCGCGCGCGCGCTGCGAGCTTCCAGCGCGCGCGCGCTTCAAGTTCCTGC 516
Db 430 CCAGCGCGCGCGCGCGCTGCGAGCTTCCAGCGCGCGCGCGCTTCAAGTTCCTGC 489
QY 517 AGTTGCTTACTTCAAGGCAATATCTCCCAAGCCATCGCGCTGCACACTTC 576
Db 490 AGTTGCTTACTTCAAGGCAATATCTCCCAAGCCATCGCGCTGCACACTTC 549
QY 577 TACTGAAGCATCTGATGACGAATGATGAAGGAACATGGCATATTATAGAGCGCTGC 636
Db 550 TACTGAAGCATCTGATGACGAATGATGAAGGAACATGGCATATTATAGAGCGCTGC 609
QY 637 CTGGTCCGAGGACTACATTAAAGACCTGGAACCAAGTCATATGAAGCCTTTCATCC 696
Db 610 CTGGTCCGAGGACTACATTAAAGACCTGGAACCAAGTCATATGAAGCCTTTCATCC 669
QY 697 GAGCAGTGGGGGATCAACGGTGAGACTGAGAACATCCATCAGACATGGAGCTGG 756
Db 670 GAGCAGTGGGGGATCAACGGTGAGACTGAGAACATCCATCAGACATGGAGCTGG 729
QY 757 CCGTCCGAGCTTCTCAAGCGCTTTACGAGTGTCTCGACCGCTCGAGGCTTCCAGG 816
Db 730 CCGTCCGAGCTTCTCAAGCGCTTTACGAGTGTCTCGACCGCTCGAGGCTTCCAGG 789
QY 817 AGATCAAGACTTCAAGGATTTCTACCTTTCATAGCAGATCATATGATGAGTCTCG 876
Db 790 AGATCAAGACTTCAAGGATTTCTACCTTTCATAGCAGATCATATGATGAGTCTCG 849
QY 877 AATGCAAAATACAGTGTGAAGAACTCACCCAGTTATAGAGGCTATCCGTTGAGA 936
Db 850 AATGCAAAATACAGTGTGAAGAACTCACCCAGTTATAGAGGCTATCCGTTGAGA 909
QY 937 AATTGTGGTACCATGTATCATTTCTGAGTTTGCCTATTATAGTTGAAGCACTGA 996
Db 910 AATTGTGGTACCATGTATCATTTCTGAGTTTGCCTATTATAGTTGAAGCACTGA 969
QY 997 AGAATCAGCGCCCTGTGCGAGTCAGCTATCTGCTCTTTGATCAGATGACAAAGGTCTATGC 1056
Db 970 AGAATCAGCGCCCTGTGCGAGTCAGCTATCTGCTCTTTGATCAGATGACAAAGGTCTATGC 1029
QY 1057 AGCAGAACTGTGTATTACCAAGTACCAAGGACACTTTGGGCGCTCTCGGATGAGCACT 1116
Db 1030 AGCAGAACTGTGTATTACCAAGTACCAAGGACACTTTGGGCGCTCTCGGATGAGCACT 1089
QY 1117 TCCAGCCGAGCTGAGCGAGTTCAGTTCTTTATGTGACCACTCCAGAGGAGCTGT 1176
Db 1090 TCCAGCCGAGCTGAGCGAGTTCAGTTCTTTATGTGACCACTCCAGAGGAGCTGT 1149
QY 1177 ATGACTTTGCTAAGGAAATATAATGGATGATGAGGAGAAAGTTGTGGAATATGTGG 1236
Db 1150 ATGACTTTGCTAAGGAAATATAATGGATGATGAGGAGAAAGTTGTGGAATATGTGG 1209
QY 1237 ATGACTTTGGAACCTGGAGGAGACAGCTAGGCCACAGCAACCAAGAGACTTCTCTTT 1296
Db 1210 ATGACTTTGGAACCTGGAGGAGACAGCTAGGCCACAGCAACCAAGAGACTTCTCTTT 1269
QY 1297 GGGTTTCAGAAACACAGATTTCTTTCTTTTCCCAACAGCCGAGCTTTGATACCTTC 1356
Db 1270 GGGTTTCAGAAACACAGATTTCTTTCTTTTCCCAACAGCCGAGCTTTGATACCTTC 1329
QY 1357 AGAGCCTTCTCTTTACTCTCCAAAGTGAAGGGAAGCCCGGCTCTCTCTAACTGATGTC 1416
```

```
Db 1330 AGAGCCTTCTCTTTACTCTCTCAAAGTGAAGGGAAGCCCGCTCCCTCTACCTGCAGC 1389
QY 1417 ATCA 1420
Db 1390 AACA 1393

RESULT 12
NMAJ6469 1685 bp mRNA linear ROD 02-AUG-1999
LOCUS Mus musculus mRNA for cartilage-associated protein (CASP).
DEFINITION AJ006469
ACCESSION AJ006469
VERSION AJ006469.1 GI:3687319
KEYWORDS cartilage-associated protein; CASP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1
AUTHORS Morello, R., Tonachini, L., Monticone, M., Viggiano, L., Rocchi, M.,
Canceda, R., and Castagnola, P.
TITLE cDNA cloning, characterization and chromosome mapping of Crtp
encoding the mouse cartilage associated protein
JOURNAL Matrix Biol. 18 (3), 319-324 (1999)
MEDLINE 99357019
PubMed 10429950
REFERENCE 2 (bases 1 to 1685)
AUTHORS Castagnola, P.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Castagnola P., Biotechnology in Oncologia,
Ist. Naz.le per la Ricerca sul Cancro, L.go R. Benzi 10 Genoa,
16132, ITALY
FEATURES
source
1. 1685
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="embryo"
/dev_stage="14-18 d.p.c."
1. 1685
/gene="casp"
72. 1274
/gene="casp"
/codon_start=1
/product="cartilage-associated protein (CASP)"
/protein_id="CAA07053.1"
/db_xref="GI:3687320"
/db_xref="SPTREMBL:O88698"
/translaton="MGPSPAAALLVLLCVGCAPTGPGYERYSPRNFPRDEIMPLE
SAVRHALDOYSGEHWAEISVGLVSLRLHLLRDSBAFCHRNCSAATPAPAGPASH
AEURLFGSVLRRAQCLKRCQGLPAPQSPQSPSVLADFOQRPYKFLQFAYPKANDL
PKAAATAATYLLKHPDDEMKRNMYKSLFGAEDHDKOLETKSYSLFVRAVYNG
ENWRTSLSDMELAPDLKAFYECIAACGSEIRIKDFKDFYLSIADHYVEVLCKIRC
EETLTPVIGGYPVKVFATMYHYLFQAYYKINDLKNAAPCAVSLLFDQSDRVNQNL
VYQYHRDKWGLSDHEHFQPRPEAVQFNVTTLQELYDFAQELHMDDEGEVVEYVDD
LLETEESA"
72. 146
/sig_peptide
/gene="casp"
147. 1271
/mat_peptide
/gene="casp"
/product="cartilage-associated protein (CASP)"
BASE COUNT 362 a 503 c 449 g 371 t
ORIGIN

Query Match 24.4%; Score 944; DB 10; Length 1685;
Best Local Similarity 83.6%; Pred. No. 1.9e-153;
Matches 1080; Conservative 2; Mismatches 207; Indels 3; Gaps 1;
QY 2 TTCCTTCTCCCTCCCTTTTCCCTTCTCTGCTCCCTTCTCTTCTCTTCTCTTCTCGCGGCGC 61
Db 11 TTCCTAGCGCTCGCGCTCGCGCTCTCTGCTCGGCGCTCTGCTTCTGCGGCGCACGCGC 70
```









```

* 1 644: contig of 644 bp in length
* 645: gap of 100 bp
* 745: contig of 684 bp in length
* 1428: gap of 100 bp
* 1429: contig of 692 bp in length
* 1529: gap of 100 bp
* 2221: contig of 693 bp in length
* 3013: gap of 100 bp
* 3114: contig of 655 bp in length
* 3768: gap of 100 bp
* 3888: contig of 636 bp in length
* 4504: gap of 100 bp
* 4505: contig of 669 bp in length
* 5273: gap of 100 bp
* 5374: contig of 670 bp in length
* 6044: gap of 100 bp
* 6143: contig of 680 bp in length
* 6823: gap of 100 bp
* 6824: contig of 678 bp in length
* 6924: gap of 100 bp
* 7602: contig of 679 bp in length
* 7702: gap of 100 bp
* 8381: contig of 671 bp in length
* 8481: gap of 100 bp
* 9251: contig of 694 bp in length
* 9252: gap of 100 bp
* 9946: contig of 688 bp in length
* 10046: gap of 100 bp
* 10734: contig of 683 bp in length
* 10834: gap of 100 bp
* 1117: contig of 654 bp in length
* 11617: gap of 100 bp
* 12270: contig of 654 bp in length
* 12370: gap of 100 bp
* 13029: contig of 659 bp in length
* 13030: gap of 100 bp
* 13129: contig of 667 bp in length
* 13130: gap of 100 bp
* 13797: contig of 690 bp in length
* 14586: gap of 100 bp
* 14587: contig of 670 bp in length
* 14686: gap of 100 bp
* 15356: contig of 652 bp in length
* 15456: gap of 100 bp
* 15457: contig of 652 bp in length
* 16109: gap of 100 bp
* 16208: contig of 685 bp in length
* 16894: gap of 100 bp
* 16994: contig of 690 bp in length
* 17684: gap of 100 bp
* 17784: contig of 639 bp in length
* 18422: gap of 100 bp
* 18522: contig of 690 bp in length
* 19212: gap of 100 bp
* 19312: contig of 695 bp in length
* 20007: gap of 100 bp
* 20107: contig of 657 bp in length
* 20764: gap of 100 bp
* 20864: contig of 659 bp in length
* 21523: gap of 100 bp
* 21623: contig of 652 bp in length
* 22275: gap of 100 bp
* 22375: contig of 680 bp in length
* 23055: gap of 100 bp
* 23155: contig of 648 bp in length
* 23803: gap of 100 bp
* 23903: contig of 674 bp in length
* 24577: gap of 100 bp
* 24678: contig of 685 bp in length
* 25362: gap of 100 bp
* 25462: contig of 684 bp in length
* 26146: gap of 100 bp
* 26246: contig of 625 bp in length
* 26871: gap of 100 bp
* 26971: contig of 690 bp in length
* 27661: gap of 100 bp
* 27761: contig of 624 bp in length
* 28385: gap of 100 bp

```

```

28386: gap of 100 bp
28486: contig of 658 bp in length
29143: gap of 100 bp
29243: contig of 662 bp in length
29905: gap of 100 bp
30005: contig of 677 bp in length
30684: gap of 100 bp
30782: contig of 673 bp in length
31455: gap of 100 bp
31555: contig of 672 bp in length
32227: gap of 100 bp
32327: contig of 684 bp in length
33011: gap of 100 bp
33111: contig of 689 bp in length
33800: gap of 100 bp
33900: contig of 690 bp in length
34590: gap of 100 bp
34690: contig of 688 bp in length
35378: gap of 100 bp
35478: contig of 657 bp in length
36135: gap of 100 bp
36235: contig of 658 bp in length
36893: gap of 100 bp
36993: contig of 672 bp in length
37665: gap of 100 bp
37765: contig of 688 bp in length
38433: gap of 100 bp
38533: contig of 684 bp in length
39217: gap of 100 bp
39317: contig of 679 bp in length
39996: gap of 100 bp
40096: contig of 690 bp in length
40786: gap of 100 bp
40886: contig of 661 bp in length
41547: gap of 100 bp
41647: contig of 690 bp in length
42337: gap of 100 bp
42437: contig of 682 bp in length
43119: gap of 100 bp
43219: contig of 694 bp in length
43913: gap of 100 bp
44013: contig of 647 bp in length
44660: gap of 100 bp
44760: contig of 670 bp in length
45430: gap of 100 bp
45530: contig of 677 bp in length
46207: gap of 100 bp
46307: contig of 679 bp in length
46986: gap of 100 bp
47086: contig of 682 bp in length
47768: gap of 100 bp
47868: contig of 658 bp in length
48526: gap of 100 bp
48626: contig of 680 bp in length
49306: gap of 100 bp
49406: contig of 649 bp in length
50055: gap of 100 bp
50155: contig of 690 bp in length
50845: gap of 100 bp
50945: contig of 703 bp in length
51649: gap of 100 bp
51749: contig of 701 bp in length
52449: gap of 100 bp
52549: contig of 660 bp in length
53209: gap of 100 bp
53309: contig of 637 bp in length
53946: gap of 100 bp
54046: contig of 658 bp in length
54704: gap of 100 bp

```

Query Match 13.2%; Score 510.2; DB 2; Length 64706;  
 Best Local Similarity 99.4%; Pred. No. 2.2e-78;  
 Matches 520; Conservative 2; Mismatches 0; Indels 1; Gaps 1;



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 30, 2004, 07:10:40 ; Search time 4130.11 Seconds  
(without alignments)  
11232.506 Million cell updates/sec

Title: US-09-729-674-1\_COPY\_132\_1265

Perfect score: 1134

Sequence: 1 gggcgcccaatacgaacg.....tgggaactggagagaccagc 1134

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.cm.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.fod.\*

36: em.htg.mam.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1133.2	99.9	2307	9	HSAJ6470 Homo sapi
2	1133.2	99.9	3871	6	BD140451 Secreted
3	1128.4	99.5	1439	9	BC008745 Homo sapi
4	1128.4	99.5	1961	8	BD156517 Primer fo
5	1128.4	99.5	1961	9	BD156517 Homo sapi
6	1118	98.6	2010	9	AK075401 Homo sapi
7	1042.4	91.9	1922	9	AK091772 Homo sapi
8	1019.6	89.9	1903	9	AK092990 Homo sapi
9	897.2	79.1	1663	10	BC049890 Mus muscu
10	895.6	79.0	1685	10	MMAJ6459 Mus muscu
11	850	75.0	1810	9	AK097650 Homo sapi
12	587.6	51.8	818	9	BD148800 Primer fo
13	494.6	43.6	1665	5	GGCARPR
14	455.4	40.2	1745	9	BC011701 Homo sapi
15	455.4	40.2	1889	9	BC001047 Homo sapi
16	455.4	40.2	2037	9	BC007942 Homo sapi
17	455.4	40.2	2347	6	AX335490 Sequence
18	455.4	40.2	2347	9	HSU47621 Homo sapien
19	455.4	40.2	2443	6	AX714040 Sequence
20	455.4	40.2	2443	9	AK056085 Homo sapi
21	453.8	40.0	2079	9	HSA250583 Homo sapi
22	418.2	36.9	1903	10	BC031856 Mus muscu
23	406.6	35.9	182509	9	AC112211 Homo sapi
24	403	35.5	1407	10	RNSG65MR X55454 R. norvegicu
25	379.2	33.4	64706	2	AC136358 Homo sapi
26	320.8	28.3	226615	2	AC128652 Rattus no
27	193.6	17.1	178345	9	AC091172 Homo sapi
28	193.6	17.1	196735	2	AC091192 Homo sapi
29	169.2	14.9	176789	2	AC032013 Mus muscu
30	169.2	14.9	217092	10	AL590968 Mouse DNA
31	166	14.6	226822	2	AC112575 Rattus no
32	166	14.6	241607	2	AC113636 Rattus no
33	156.4	13.8	2524	6	BD156505 Primer fo
34	156.4	13.8	2524	9	AK027680 Homo sapi
35	156.4	13.8	2563	6	AX136191 Sequence
36	156.4	13.8	2563	6	BD123548 Secretary
37	156.4	13.8	2563	9	AK075418 Homo sapi
38	156.4	13.8	2583	6	AX704765 Sequence
39	156.4	13.8	2585	6	BD156581 Primer fo
40	156.4	13.8	2585	9	AK027697 Homo sapi
41	156.4	13.8	2993	6	BD158280 Primer fo
42	156.4	13.8	2993	9	AK027648 Homo sapi
43	154.8	13.7	2600	9	AF097432 Homo sapi
44	154.8	13.7	2829	9	AF097431 Homo sapi
45	147.4	13.0	2249	10	MMU430350 Mus muscu

# ALIGNMENTS

RESULT 1  
HSAJ6470 HSAJ6470 2307 bp mRNA linear PRI 23-MAR-2000  
LOCUS Homo sapiens mRNA for cartilage-associated protein (CASP).  
DEFINITION  
ACCESSION AJ006470  
VERSION AJ006470.1 GI:3687321  
KEYWORDS cartilage-associated protein; CASP.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS

1  
Tonachini L., Morello R., Monticone M., Skaug J., Scherer S.W.,  
Cancedda R. and Castagnola P.  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CDNA cloning, characterization and chromosome mapping of the gene

```
JOURNAL encoding human cartilage associated protein (CRTAP)
MEDLINE Cytogenet. Cell Genet. 87 (3-4), 191-194 (1999)
20169181
PUBMED 10702564
REFERENCE 2 (bases 1 to 2307)
AUTHORS Castagnola, P.
DIRECT SUBMISSION
TITLE Submitted (12-JUN-1998) Castagnola P., Biotechnology in Oncologia,
JOURNAL Ist. Naz.le per la Ricerca sul Cancro, L.go R. Benzi 10 Genoa,
16132, ITALY
FEATURES
    source Location/Qualifiers
        1..2307
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="Contig of pCPH2 and I.M.A.G.E. cloneID 1090695"
            /tissue_type="fetal brain"
            /dev_stage="19-23 weeks-Old"
        1..2307
            /gene="casp"
            12..1217
            /gene="casp"
            /gene="casp"
            /codon_start=1
            /product="cartilage-associated protein (CASP)"
            /protein_id="CAA07054.1"
            /db_xref="GI:3687322"
            /translation="MEFGREGAALALLVACALRAGRAQYRYFRSFRPDELMLPL
            ESAYHALDKVSGHSAESVYLEISLRHLRLDSEAFCHRNCSAOPPEPAAGLAS
            YPELRFGGLRLAKHCKRCKQFQSPQREVLADFORPEPYKFLQFAVFKANN
            LPKAIAAAHFLAKHDPDEMKRNNAVYKSLPGAEDVIKDLTKSVESLFIKAVRAN
            GENWRTSITDMELALPFPKAFVECLAACSGSREIKDPKDFYLSIADHYVELEKIQ
            CEENLTPVIGSFVPEKFAVTMIHYLQAFATKINDLKNAPCAVSYLLFDNDKWMQON
            LVYYQHRDWTGLSDEHFQRPPEAVQFNVTTLQKELYDFAKENIMDDDEGEVVEYVD
            DLELEETS"
        12..89
            /sig_peptide
        90..1214
            /mat_peptide
            /gene="casp"
            /product="cartilage-associated protein (CASP)"
BASE COUNT 545 a 533 c 548 g 578 t 3 others
ORIGIN
Query Match 99.9%; Score 1133.2; DB 9; Length 2307;
Best Local Similarity 99.8%; Pred No. 2.4e-180;
Matches 1132; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGGCGGCCCAATACGAACGCTACAGCTTCGCGAGCTTCCACGGGACGAGCTGATGCCG 60
DB 81 GGGGCGGCCCAATACGAACGCTACAGCTTCGCGAGCTTCCACGGGACGAGCTGATGCCG 140
QY 61 CTCGAGTCGGCTACCGGACGCGCTGGACAGTACAGCGCGGCGAGCACTGGCGCGAGAGC 120
DB 141 CTCGAGTCGGCTACCGGACGCGCTGGACAGTACAGCGCGGCGAGCACTGGCGCGAGAGC 200
QY 121 STKGCTACCTGGAGATCAGCTGCGGCTGCACCGCTTGTGCGCGACGAGCGGCTTC 180
DB 201 GTGGGCTACCTGGAGATCAGCTGCGGCTGCACCGCTTGTGCGCGACGAGCGGCTTC 260
QY 181 TGCCACCGCACTGAGCGCGCGCGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 240
DB 261 TGCCACCGCACTGAGCGCGCGCGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 320
QY 241 CCGGAGCTGCGCTCTTCTGGGGGCTGTGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 300
DB 321 CCGGAGCTGCGCTCTTCTGGGGGCTGTGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCTGCAAG 380
QY 301 CAGGCGCTGCGAGCTTCCGCGAGTCCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCCAG 360
DB 381 CAGGCGCTGCGAGCTTCCGCGAGTCCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCCAG 440
QY 361 CGCGCGGCGGCGGCTTCCGCGAGTCCGAGCTTCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCCAG 420
```



PR 26-NOV-1997 US 60/066804.23-NOV-1998 US 09/197886 PI  
KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A COLLINS PI  
PI CHERYL EVANS, DAVID MERBERG, MAURICE TREACY, MICHAEL J AGOSTINO,  
PI ROBERT J STEININGER II, GORDON G WONG, HILARY F CLARK, KIM PI  
FECHTEL  
PC C12N15/09, C07K14/00, C12N1/21, C12N5/10, C12P19/34, C12P21/02, PC  
C12Q1/68//  
PC A61P29/00, A61P35/00, A61P37/04, A61P37/06, C12N15/00, C12N5/00 CC  
Secreted proteins and polynucleotides encoding them FH Key  
Location/Qualifiers  
FT source 1..3871  
FT Location/Qualifiers  
1..3871  
/organism="Homo sapiens (human)"  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 965 a 913 c 1033 g 958 t 2 others  
ORIGIN  
Query Match 99.9%; Score 1133.2; DB 6; Length 3871;  
Best Local Similarity 100.0%; Pred. No. 2.4e-180;  
Matches 1134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GGGCGCGCCCAATACGAACGCTACAGCTTCGGGAGCTTCCACGGGACGAGCTGATGCCG 60  
Db 132 GGGCGCGCCCAATACGAACGCTACAGCTTCGGGAGCTTCCACGGGACGAGCTGATGCCG 191  
Qy 61 CTCGAGCTCGGCTACCGCGCAGCGCTGACAAAGTACAGCGCGGAGCACTGGGCGGAGAGC 120  
Db 192 CTCGAGCTCGGCTACCGCGCAGCGCTGACAAAGTACAGCGCGGAGCACTGGGCGGAGAGC 251  
Qy 121 STKGCTACCTGGAGATCAGCTTCGGGCTCGACCGCTTGTGCGCGACACGAGGCGCTTC 180  
Db 252 STKGCTACCTGGAGATCAGCTTCGGGCTCGACCGCTTGTGCGCGACACGAGGCGCTTC 311  
Qy 181 TGCCACCGCACTGACGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAT 240  
Db 312 TGCCACCGCACTGACGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAT 371  
Qy 241 CCCGAGCTGGCGCTTCCTCGGGGGCGCTGCTGCGCGCGCGCGCACTGCTCAAGCGCTGCAAG 300  
Db 372 CCCGAGCTGGCGCTTCCTCGGGGGCGCTGCTGCGCGCGCGCGCACTGCTCAAGCGCTGCAAG 431  
Qy 301 CAGGCGCTGCGAGCTTCGCGCAGTCCGACCGCGCGCGCGAGGTGCTGGCGGACTTCCAG 360  
Db 432 CAGGCGCTGCGAGCTTCGCGCAGTCCGACCGCGCGCGCGAGGTGCTGGCGGACTTCCAG 491  
Qy 361 CGCGCGAGCGCTACAAAGTTCCTGCGCTGCTGCTTCAAGCAAAATATCTCCGCAAA 420  
Db 492 CGCGCGAGCGCTACAAAGTTCCTGCGCTGCTGCTTCAAGCAAAATATCTCCGCAAA 551  
Qy 421 GCCATCGCGCTGCTCAACCTTTCTACTGAAGCATCTCTGATGACGAAATGATGAAGAGG 480  
Db 552 GCCATCGCGCTGCTCAACCTTTCTACTGAAGCATCTCTGATGACGAAATGATGAAGAGG 611  
Qy 481 AACATGCATATTATAGAGCTGCTGCTGCGCGAGCACTACATTAAGACCTGGAACC 540  
Db 612 AACATGCATATTATAGAGCTGCTGCTGCGCGAGCACTACATTAAGACCTGGAACC 671  
Qy 541 AAGTCATATGAAGCGCTGTTTCATCCGAGCAGTGGCGGCATACACCGTGAAGACTGGAGA 600  
Db 672 AAGTCATATGAAGCGCTGTTTCATCCGAGCAGTGGCGGCATACACCGTGAAGACTGGAGA 731  
Qy 601 ACATCCATCAGACATGGAGCTGGCGCTTCCGAGCTTCTTCAAGCGCTTTTACGAGTGT 660  
Db 732 ACATCCATCAGACATGGAGCTGGCGCTTCCGAGCTTCTTCAAGCGCTTTTACGAGTGT 791  
Qy 661 CTCGAGCGCTCGAGGGTTCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATA 720  
Db 792 CTCGAGCGCTCGAGGGTTCAGGAGATCAAGGACTTCAAGGATTTCTACCTTTCCATA 851

QY 721 GCAGATCATTTATGTAGAAAGTTCTTGGATGCAAAATACAGTGTGAAGAGAACTCACCCCA 780  
DB 852 GCAGATCATTTATGTAGAAAGTTCTTGGATGCAAAATACAGTGTGAAGAGAACTCACCCCA 911  
QY 781 GTTATAGAGGCTATCCCGTTGAGAAATTTGGCTACCATGTATCATCTTGCAGTTT 840  
DB 912 GTTATAGAGGCTATCCCGTTGAGAAATTTGGCTACCATGTATCATCTTGCAGTTT 971  
QY 841 GCCTATTATAGTTCGAAACGACTGAAAGTATGAGCCCTGTGCACTGAGTATCTGCTC 900  
DB 972 GCCTATTATAGTTCGAAACGACTGAAAGTATGAGCCCTGTGCACTGAGTATCTGCTC 1031  
QY 901 TTTGATCAGATGACAAGGTCATGACGAGCAACCTGTGTGTATTTACCATACACAGGAC 960  
DB 1032 TTTGATCAGATGACAAGGTCATGACGAGCAACCTGTGTGTATTTACCATACACAGGAC 1091  
QY 961 ACTTGGGGCTCTCGGATGAGCACTTCCAGCCAGCACTGAGCACTTCTTCTTTTAAAT 1020  
DB 1092 ACTTGGGGCTCTCGGATGAGCACTTCCAGCCAGCACTGAGCACTTCTTCTTTTAAAT 1151  
QY 1021 GTGACCACACTCCAGAAGGAGCTGTATGACTTTGCTAAGGAAATATATATGATGATGAT 1080  
DB 1152 GTGACCACACTCCAGAAGGAGCTGTATGACTTTGCTAAGGAAATATATATGATGATGAT 1211  
QY 1081 GAGGAGAGAGTTGCTGAATATGAGTACCTCTTGGAACTGGAGGAGCAGC 1134  
DB 1212 GAGGAGAGAGTTGCTGAATATGAGTACCTCTTGGAACTGGAGGAGCAGC 1265  
RESULT 3  
BC008745  
LOCUS  
DEFINITION  
Homo sapiens, cartilage associated protein, clone MGC:1926  
IMAGE:3347384, mRNA, complete cds.  
ACCESSION  
BC008745  
VERSION  
BC008745.1 GI:14250580  
KEYWORDS  
MGC.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1439)  
AUTHORS  
Straussberg, R.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (25-MAY-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/DTP  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@hgrl.nih.gov](mailto:nisc_mgc@hgrl.nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantropop, S., Thomas, P.J.,  
Tingerson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,  
Zhang, J.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IMAGE Plate: 5 Row: 1 Column: 24  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 3687321.  
Location/Qualifiers

```
source
1. 1439
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LOCUSID:10491"
/db_xref="taxon:9606"
/clone="MGC:1926 IMAGE:3347384"
/tissue_type="kidney, renal cell adenocarcinoma"
/clone_lib="NIH_MGC_14"
/lab_host="DH10B-R"
/notes="vector: pOTB7"
36. 1241
/codon_start=1
/product="cartilage associated protein"
/protein_id="AAH08745.1"
/db_xref="GI:14250581"
/translation="MPEGRGMAALLLVCALRAGRAQYVRSFPRDELMLPL
ESAYRALDKYSGHWAESVGLLEISLRLLHLLRDSFAFCHRNCSAPPOPEAAGLAS
YPERLFGLLRRARHCKRCKGFLPFRQPSRDVLADQRRPEPKFLQFAYFRANN
LPKAAAHATFLKHPDDEMMKNMAYIKLPGAEDYIKOLETKSYHESULFIRAVRAYN
GENWRTITMELALPFFKFAFYELCAACGSREIKDFKFIYSIADHYFVLECKIQ
CEENLPTVIGGYPYKVFVAMHYVLOFAYVYKLNDAKNAAPCAVSYLLFDQNDKVMQON
LVYQYHRDHWGLSDEHFQPRPEAVQFFNVTTLKELYDFAKENIMDDDEGEVVEVD
DLLELETS"
BASE COUNT 362 a 415 c 375 g 287 t
ORIGIN
Query Match 99.5%; Score 1128.4; DB 9; Length 1439;
Best Local Similarity 99.6%; Pred. No. 1.6e-179;
Matches 1129; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 GGGCGCGCCCAATACGACGCTACAGCTTCGCGAGCTTCCACGGGACGAGCTGATGCCG 60
DB 105 GGGCGCGCCCAATACGACGCTACAGCTTCGCGAGCTTCCACGGGACGAGCTGATGCCG 164
QY 61 CTCGAGTCGGCTACCGGACGCGCTGGACAAGTACAGCGCGGACGACTGGCGCGAGAGC 120
DB 165 CTCGAGTCGGCTACCGGACGCGCTGGACAAGTACAGCGCGGACGACTGGCGCGAGAGC 224
QY 121 STKGCTACCTGGAGATCAGCTTCGGGCTGACCGCTTGTGCGCGGACGAGGCGCTTC 180
DB 225 GTGGCTACCTGGAGATCAGCTTCGGGCTGACCGCTTGTGCGCGGACGAGGCGCTTC 284
QY 181 TGCCACCGCAACTGACGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAT 240
DB 285 TGCCACCGCAACTGACGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAT 344
QY 241 CCGGAGCTGCGCTTCTCGGGGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAAG 300
DB 345 CCGGAGCTGCGCTTCTCGGGGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAAG 404
QY 301 CAGGCGCTGCGAGCTTCGGCGAGTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCAG 360
DB 405 CAGGCGCTGCGAGCTTCGGCGAGTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCAG 464
QY 361 CGCGCGAGCGCTTACAAGTTCCTGAGTTCGCTTACTTCAAGGCAAAATATCTCCCAAA 420
DB 465 CGCGCGAGCGCTTACAAGTTCCTGAGTTCGCTTACTTCAAGGCAAAATATCTCCCAAA 524
QY 421 GCATCGCGCTGCTCAGACCTTCTGAGCGTCTGAGCGTCTGAGCGTCTGAGCGTCTGAGCG 480
DB 525 GCATCGCGCTGCTCAGACCTTCTGAGCGTCTGAGCGTCTGAGCGTCTGAGCGTCTGAGCG 584
QY 481 AACATGGCATATTATAAGAGCTGCTGCTGCGGAGGACTACATTAAGAGCTGGAAC 540
DB 585 AACATGGCATATTATAAGAGCTGCTGCTGCGGAGGACTACATTAAGAGCTGGAAC 644
QY 541 AAGTCATATGAAGCTGTTCATCCGAGCGTGGGCGATCAACCGGTGAGAACTGGAGA 600
DB 645 AAGTCATATGAAGCTGTTCATCCGAGCGTGGGCGATCAACCGGTGAGAACTGGAGA 704
QY 601 ACATCCATCAGAGATGGGCTGCGCTTCCGAGCTTCTTCAAGCGCTTTTACGAGTGT 660
DB 705 ACATCCATCAGAGATGGGCTGCGCTTCCGAGCTTCTTCAAGCGCTTTTACGAGTGT 764
```



```
Db 289 TGCCACGCAACTGACGCGCGCGCGAGCCGCGCGCTCGCAGCTAT 348
Qy 241 CCCGAGCTGCGCTCTTCGCGGGGCTGCTGCGCGCGCGCGCTCGCTCAAGCGCTGGAAG 300
Db 349 CCCGAGCTGCGCTCTTCGCGGGGCTGCTGCGCGCGCGCGCTCGCTCAAGCGCTGGAAG 408
Qy 301 CAGGGCTGCGCGCTTCGCGCACTCCGCGAGCCGCGCGCGAGGTGCTGGCGGACTTCCAG 360
Db 409 CAGGGCTGCGCGCTTCGCGCACTCCGCGAGCCGCGCGCGAGGTGCTGGCGGACTTCCAG 468
Qy 361 CGCGCGAGCCCTACAAGTTCCTGAGTTCGCTTACGTTACGTTACGTTACGTTACGTTACGTT 420
Db 469 CGCGCGAGCCCTACAAGTTCCTGAGTTCGCTTACGTTACGTTACGTTACGTTACGTTACGTT 528
Qy 421 GCCATCGCGCTGCTCACACCTTCTACTGAGCATCCTGATGACGAAATGATGAGAGG 480
Db 529 GCCATCGCGCTGCTCACACCTTCTACTGAGCATCCTGATGACGAAATGATGAGAGG 588
Qy 481 AACATGCGATATTATAAGAGCTCGCTGCTGCGGAGCTACATTTAAAGACCTGGAAACC 540
Db 589 AACATGCGATATTATAAGAGCTCGCTGCTGCGGAGCTACATTTAAAGACCTGGAAACC 648
Qy 541 AAGTCATATGAAGCTGTTTCATCGAGCAGTGGCGCATACACGCTGAGACTGCGA 600
Db 649 AAGTCATATGAAGCTGTTTCATCGAGCAGTGGCGCATACACGCTGAGACTGCGA 708
Qy 601 ACATCCCATCAGACATGAGCTGCGCTGCTGCGGAGCTTCTTCAAGGCTTTTACGAGTGT 660
Db 709 ACATCCCATCAGACATGAGCTGCGCTGCTGCGGAGCTTCTTCAAGGCTTTTACGAGTGT 768
Qy 661 CTCGAGCTGCGAGGTTCCAGGAGATCAGAGCTTCAAGATTTCTACCTTTCCATA 720
Db 769 CTCGAGCTGCGAGGTTCCAGGAGATCAGAGCTTCAAGATTTCTACCTTTCCATA 828
Qy 721 CGAGATCATTATGAGAGTTCTGAAATGCAAAATACAGTGTGAAGAGAACTCACCCCA 780
Db 829 CGAGATCATTATGAGAGTTCTGAAATGCAAAATACAGTGTGAAGAGAACTCACCCCA 888
Qy 781 GTTATGAGAGCTATCCGGTTGAGAAATTTGTGCTACCATGATATCATTTCTGAGTTT 840
Db 889 GTTATGAGAGCTATCCGGTTGAGAAATTTGTGCTACCATGATATCATTTCTGAGTTT 948
Qy 841 GCCTATTATAGTGAAGCTGAGAAATGCAAGCTGCGGCTGCGAGTCACTATCTGCTC 900
Db 949 GCCTATTATAGTGAAGCTGAGAAATGCAAGCTGCGGCTGCGAGTCACTATCTGCTC 1008
Qy 901 TTTGATCAGAAATGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db 1009 TTTGATCAGAAATGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1068
Qy 961 ACTTGGGCTCTCGGATGAGACTTCCAGGCGAGCTGAGAGTTCAGTTCCTTTAAT 1020
Db 1069 AGTGGGCTCTCGGATGAGACTTCCAGGCGAGCTGAGAGTTCAGTTCCTTTAAT 1128
Qy 1021 GTGACCACTCCAGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
Db 1129 GTGACCACTCCAGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1188
Qy 1081 GAGGAGAGTGTGGAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1134
Db 1189 GAGGAGAGTGTGGAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1242
```

## RESULT 6

```
AK075401
LOCUS AK075401 2010 bp mRNA linear PRI 03-SEP-2002
DEFINITION Homo sapiens cDNA PSEC0091 fls, clone PLACE1001683, highly similar
to Homo sapiens mRNA for cartilage-associated protein.
ACCESSION AK075401
VERSION AK075401.1 GI:22761465
KEYWORDS oligo capping, fls (full insert sequence).
SOURCE Homo sapiens (human)
```

```
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Ota, T., Nishikawa, T., Suzuki, Y., Kawai-Hio, Y., Hayashi, K.,
JOURNAL Ishii, S., Saito, K., Yamamoto, J., Wakamatsu, A., Nagai, T.,
REFERENCE Nakamura, Y., Sato, H., Nagahara, K., Sugano, S. and Isogai, T.
AUTHORS HRI human cDNA sequencing project
TITLE Unpublished
JOURNAL HRI human cDNA sequencing project
REFERENCE 2 (bases 1 to 2010)
AUTHORS Isogai, T. and Yamamoto, J.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
COMMENT Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
HRI human cDNA sequencing project; cDNA 5' - 3' - end one pass
sequencing, clone selection and full insert sequencing; Helix
Research Institute (supported by Japan Key Technology Center etc.);
cDNA library construction; Institute of Medical Science, University
of Tokyo, Laboratory of Genome Structure, Human Genome Center.
FEATURES
Location/Qualifiers
1..2010
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1001683"
/tissue_type="placenta"
/clone_lib="PLACE1"
/note="Cloning vector: pME18SFL3"
misc_difference 728*729
/note="compared to AK001634 and AJ006470"
/replace="c"
BASE COUNT 416 a 606 c 485 g 503 t
ORIGIN
Query Match 98.6%; Score 1118; DB 9; Length 2010;
Best Local Similarity 99.6%; Pred. No. 8.7e-178;
Matches 1129; Conservative 2; Mismatches 2; Indels 1; Gaps 1;
Qy 1 GGGCGCGCCCAATACGAAGCTTACAGCTTCGCGAGCTTCCACGGAGAGCTGATGCGG 60
Db 165 GGGCGCGCCCAATACGAAGCTTACAGCTTCGCGAGCTTCCACGGAGAGCTGATGCGG 224
Qy 61 CTCGAGTGGCGCTACCGGACGCGCTGGACAAGTACAGCGCGAGCACTGGCGCGAGAGC 120
Db 225 CTCGAGTGGCGCTACCGGACGCGCTGGACAAGTACAGCGCGAGCACTGGCGCGAGAGC 284
Qy 121 STKGCTACCTGGAGATCAGCTTCGCGGTGCAACCGCTTGTCTGCGGAGCAGCGAGCTTC 180
Db 285 GTGGGCTACCTGGAGATCAGCTTCGCGGTGCAACCGCTTGTCTGCGGAGCAGCGAGCTTC 344
Qy 181 TGCCACCGCAACTGCGAGCGCGCGCGAGCGCGCGAGCGCGCGCTGCCAGCTAT 240
Db 345 TGCCACCGCAACTGCGAGCGCGCGCGCGAGCGCGCGAGCGCGCGCTGCCAGCTAT 404
Qy 241 CCCGAGCTCGCGCTCTTTCGGGGGCTGCTGCGCGCGCGCTGCTGCGCGCGCTGCAAG 300
Db 405 CCCGAGCTCGCGCTCTTTCGGGGGCTGCTGCGCGCGCGCTGCTGCGCGCGCTGCAAG 464
Qy 301 CAGGGCTGCGAGCTTCGCGAGTCCAGCCAGCCAGCCAGCGAGGTGCTGGCGAGCTTCAG 360
Db 465 CAGGGCTGCGAGCTTCGCGAGTCCAGCCAGTCCAGCCAGCGAGGTGCTGGCGAGCTTCAG 524
Qy 361 CGCGCGAGCGCTTACAAGTTCCTGAGTTCGCTTACTTCAAGGCAAAATATCTCCCAAA 420
Db 525 CGCGCGAGCGCTTACAAGTTCCTGAGTTCGCTTACTTCAAGGCAAAATATCTCCCAAA 584
Qy 421 GCATCGCGCTGCTCACACCTTCTTCTGAGAGCATCTGATGAGCAAAATGATGAGAGG 480
Db 585 GCATCGCGCTGCTCACACCTTCTTCTGAGAGCATCTGATGATGATGATGATGATGAGAGG 644
Qy 481 AACATGCGATATTATAAGAGCTGCTGCTGCGGAGCTACATTTAAAGACCTGGAAACC 540
```

Db 645 AACATGGCATATTATTAAGAGCCTGCTGCTGGTCCGAGGACTACATTAAAGACCTGGAACC 704  
QY 541 AGTCTATTAAGAGCCTGTTTATCCAGCAGTCCGGGCATACAAAGCTGAGAACTGGAGA 600  
Db 705 AAGTCATATGAAGCCTGTTTAT-CCAGCAGTCCGGGCATACAAAGCTGAGAACTGGAGA 763  
QY 601 ACATCCATCACAGACATGAGCTGGCCCTTCCGACTTCTTCAAGCCCTTTTACAGTGT 660  
Db 764 ACATCCATCACAGACATGAGCTGGCCCTTCCGACTTCTTCAAGCCCTTTTACAGTGT 823  
QY 661 CTGGAGCCTGAGAGGTTTCCAGGAGATCAAGGACTTCAAGGATTTTACCTTTCCATA 720  
Db 824 CTGGAGCCTGAGAGGTTTCCAGGAGATCAAGGACTTCAAGGATTTTACCTTTCCATA 883  
QY 721 GCAGATCATATGTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAACTCCACCCCA 780  
Db 884 GCAGATCATATGTAGAGTTCTGGAATGCAAAATACAGTGTGAAGAACTCCACCCCA 943  
QY 781 GTTATAGGAGGCTATCCGGTTGAGAAATTTGGCTTACCATGATCATTTACCTTGAGTGT 840  
Db 944 GTTATAGGAGGCTATCCGGTTGAGAAATTTGGCTTACCATGATCATTTACCTTGAGTGT 1003  
QY 841 GCTATATTAAGTTGAACGCTGAGAAATGAGCCCTGTCAGTCAAGTATCTGCTC 900  
Db 1004 GCTATATTAAGTTGAACGCTGAGAAATGAGCCCTGTCAGTCAAGTATCTGCTC 1063  
QY 901 TTTGATCAGAAATGACAGGTTTCATGAGCAGAACTGCTGTGTTATACAGTACCACAGGAC 960  
Db 1064 TTTGATCAGAAATGACAGGTTTCATGAGCAGAACTGCTGTGTTATACAGTACCACAGGAC 1123  
QY 961 ACTTGGGGCTCTCGGATGAGCACTTCCAGCCAGACCTGAGCAGTTCAGTTCITTAAT 1020  
Db 1124 ACTTGGGGCTCTCGGATGAGCACTTCCAGCCAGACCTGAGCAGTTCAGTTCITTAAT 1183  
QY 1021 GTGACCACTCTCCAGAGGAGCTGTATGACTTTGCTAAGAAATATAATGATGATGAT 1080  
Db 1184 GTGACCACTCTCCAGAGGAGCTGTATGACTTTGCTAAGAAATATAATGATGATGAT 1243  
QY 1081 GAGGAGAAATGTTGGAAATGTTGATGAGCTCTTGGAACTGGAGAGACCAAC 1134  
Db 1244 GAGGAGAAATGTTGGAAATGTTGATGAGCTCTTGGAACTGGAGAGACCAAC 1297

## RESULT 7

AK091772 1922 bp mRNA linear PRI 15-JUL-2002  
LOCUS Homo sapiens cDNA FLJ34453 fis, clone HLUNG2002429, highly similar  
DEFINITION to Homo sapiens mRNA for cartilage-associated protein.

AK091772  
ACCESSION AK091772.1 GI:21750221

KEYWORDS oligo capping, fis (full insert sequence).

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S.,

Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R.,

Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,

Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,

Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,

Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A.,

Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,

Negahari, K., Masuho, Y., Nagai, K. and Isogai, T.

NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 1922)

Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7

Kazuo-Kamatari, Kisarazu, Chiba 292-0812, Japan

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.

## FEATURES

## source

1. 1922

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/c\_lone="HLUNG2002429"

/tissue\_type="lung"

/clone\_lib="HLUNG2"

/notes="cloning vector: pME18SFL3"

40. 1206

/note="unnamed protein product"

/codon\_start=1

/protein\_id="BAC03743.1"

/db\_xref="GI:21750222"

/translation="MEFGRGAALLALLCVACALRAGRAQYERYFRSPRDELMP  
ESAYRHALDKYSGEHWAEVSVDSEAFCHRNCSAOPPEAAGLASPELFGGLLR  
AHCLRKCKQGLPAFROSQPSREVLADFQREPKFLQFAYKANNLPKAAATFLL  
KHPDEMKNMAYKSLPAGAEYIKDLEKSYESLIFRAVRAYNGENMETITDMEL  
ALDPFFKAFVECLAACGSEIKDFDKFYLSDHVEVLECKICBENETPVIGYVP  
VERFVATWYHYLQFAYKLNLDKNAAPCAVSYLLFOONDKVMOONLVYVYHRDTWEL  
SDEHFQRPRAVQFFNVTTIQKELYINDDDEGEVEVEYVDDLLEETS"

BASE COUNT 413 a 566 c 467 g 476 t

## ORIGIN

Query Match 91.9%; Score 1042.4; DB 9; Length 1922;

Best Local Similarity 96.3%; Pred. No. 4.2e-165;

Matches 1092; Conservative 1; Mismatches 2; Indels 39; Gaps 1;

QY 1 GGGCGCGCCCAATACGAACGCTCAGCTTCGGAGTCTCCACGGGACGAGCTGATCGG 60  
Db 109 GGGCGCGCCCAATACGAACGCTCAGCTTCGGAGTCTCCACGGGACGAGCTGATCGG 168  
QY 61 CTCGAGTCCGCTACCGCAGCGCTCGACAAGTACAGCGCGGAGCACTGGCGCGAGAGC 120  
Db 169 CTCGAGTCCGCTACCGCAGCGCTCGACAAGTACAGCGCGGAGCACTGGCGCGAGAGC 228  
QY 121 STKGGCTACCTGGAGATCAGCTTCGGGTGCGCCGCTTGCTTCGCGCAGCAGAGCGCTTC 180  
Db 229 G-----TGCGCAGCAGCGAGCGCTTC 249  
QY 181 TGCCACCGCAACTGCGAGCGCGCGCGCAGCGCGCGCGCGCGCTCGCCAGCTAT 240  
Db 250 TGCCACCGCAACTGCGAGCGCGCGCGCAGCGCGCGCGCGCGCTCGCCAGCTAT 309  
QY 241 CCGGAGCTCGGCTCTTCGGGGGCTGCTGCGCGCGCGCGCTGCTCAAGCGTTCGCAAG 300  
Db 310 CCGGAGCTCGGCTCTTCGGGGGCTGCTGCGCGCGCGCGCTGCTCAAGCGTTCGCAAG 369  
QY 301 CAGGCGCTCGCAGCGCTTCGGCGAGTCCAGCGCGCGCGCGGAGTGTGGCGGACTTCAG 360  
Db 370 CAGGCGCTCGCAGCGCTTCGGCGAGTCCAGCGCGCGCGGAGTGTGGCGGACTTCAG 429  
QY 361 CGCGCGGAGCGCTTCAAGTCTCTGCGAGTTCGTTACTTCAAGGCAATAATCTCCCCAAA 420  
Db 430 CGCGCGGAGCGCTTCAAGTCTCTGCGAGTTCGTTACTTCAAGGCAATAATCTCCCCAAA 489  
QY 421 GCCATCGCGCTGCTCACACCTTTTCTACTGAAGCATCTCTGATGAGCAATGATGAAGAGG 480  
Db 490 GCCATCGCGCTGCTCACACCTTTTCTACTGAAGCATCTCTGATGAGCAATGATGAAGAGG 549  
QY 481 AACATGGCATATTATTAAGAGCCTGCTGGTCCGAGGACTACATTAAAGACCTGGAACC 540  
Db 550 AACATGGCATATTATTAAGAGCCTGCTGGTCCGAGGACTACATTAAAGACCTGGAACC 609  
QY 541 AAGTCATATGAAGCCTGTTTATCCAGCAGTCCGGGCATACAAAGCTGAGAACTGGAGA 600

Db 610 AAGTCATATGAAGCCCTGTTCTCCGACGAGTGGCGGCATACACGGTGAGAACTGGAGA 669  
 Qy 601 ACATCCATCACAGACATGAGCTGGCCCTCCCGACTTCTTCAAGGCTTTTACGAGTGT 660  
 Db 670 ACATCCATCACAGACATGAGCTGGCCCTCCCGACTTCTTCAAGGCTTTTACGAGTGT 729  
 Qy 661 CTCGACGCTGGAGGGTTCAGGAGATCAGGACTTCAAGGATTTCTACCTTTCCATA 720  
 Db 730 CTCGACGCTGGAGGGTTCAGGAGATCAGGACTTCAAGGATTTCTACCTTTCCATA 789  
 Qy 721 CGAGATCATATGTAGAACTTGTGAATGCAAAATACAGTGTGAGAGAACTTCAACCCA 780  
 Db 790 GCAGATCATATGTAGAACTTGTGAATGCAAAATACAGTGTGAGAGAACTTCAACCCA 849  
 Qy 781 GTTATAGGAGGCTATCCGCTTGCAGAAATTTGGGTACCATGTATCTATCTTCAGTTT 840  
 Db 850 GTTATAGGAGGCTATCCGCTTGCAGAAATTTGGGTACCATGTATCTATCTTCAGTTT 909  
 Qy 841 GCCTATTATAGTTGAACCTGGAAGAAATGCGACCCCTGTGAGTTCAGTATCTCTC 900  
 Db 910 GCCTATTATAGTTGAACCTGGAAGAAATGCGACCCCTGTGAGTTCAGTATCTCTC 969  
 Qy 901 TTGTATCAGAAATGACAAAGTCTGACGAGAACCTGGTGTATTACCATGCCACAGGAC 960  
 Db 970 TTGTATCAGAAATGACAAAGTCTGACGAGAACCTGGTGTATTACCATGCCACAGGAC 1029  
 Qy 961 ACTTGGGGCTTCGATGAGCACTTCAGCCAGACCTGAAGGAGTTCAGTTCTTTAAT 1020  
 Db 1030 ACTTGGGGCTTCGATGAGCACTTCAGCCAGACCTGAAGGAGTTCAGTTCTTTAAT 1089  
 Qy 1021 GTGACACACTCCAGAAAGAGCTGTATGACTTTGTGAAGAAATATATATGATGATGAT 1080  
 Db 1090 GTGACACACTCCAGAAAGAGCTGTATGACTTTGTGAAGAAATATATATGATGATGAT 1149  
 Qy 1081 GAGGAGAAATGTGAAATATGTGATGACCTTTTGAAGAACTGGAGGAGACGAGC 1134  
 Db 1150 GAGGAGAAATGTGAAATATGTGATGACCTTTTGAAGAACTGGAGGAGACGAGC 1203

RESULT 8  
 AK092990 1903 bp mRNA linear PRI 15-JUL-2002  
 Locus Homo sapiens cDNA FLJ35671 fis, clone SPLEN2018180, highly similar  
 DEFINITION to Homo sapiens mRNA for cartilage-associated protein.  
 AK092990  
 ACCESSION AK092990.1 GI:21751708  
 VERSION oligo capping; fis (full insert sequence).  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,  
 Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,  
 Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J.,  
 Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K.,  
 Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H.,  
 Kanda,K., Nagatsuma,M., Murakawa,K., Kanehori,K.,  
 Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,  
 Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 1903)  
 Isogai,T. and Yamamoto,J.  
 Direct Submission  
 TITLE Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
 Kazuo-Kamatori, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); cDNA library  
 construction: Helix Research Institute (HRI) (supported by Japan  
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,

HRI, and Biotechnology Center, National Institute of Technology and  
 Evaluation; clone selection for full insert sequencing: HRI and  
 RAB; annotation: HRI and RAB.  
 FEATURES  
 Location/Qualifiers  
 1. 1903  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="SPLEN2018180"  
 /tissue\_type="spleen"  
 /clone\_lib="SPLEN2"  
 /note="cloning vector: pME183FL3"  
 BASE COUNT 421 a 553 c 460 g 469 t  
 ORIGIN  
 Query Match 89.9%; Score 1019.6; DB 9; Length 1903;  
 Best Local Similarity 99.7%; Pred. No. 2.8e-161;  
 Matches 1018; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Qy 114 CGAGAGCSTKGGCTACCTGGAGATCAGCTTGGCGCTGCACCGTTGCTGCGGACAGCA 173  
 Db 147 CGAGAGCSTKGGCTACCTGGAGATCAGCTTGGCGCTGCACCGTTGCTGCGGACAGCA 206  
 Qy 174 GGCCTTCTCCACCGCAACTGCAGCGCGCGCGCGAGCCCGCGCGCGCGCTCGC 233  
 Db 207 GGCCTTCTCCACCGCAACTGCAGCGCGCGCGCGAGCCCGCGCGCGCGCTCGC 266  
 Qy 234 CAGCTATCCGAGCTGGCGCTCTTTCGGGGGCGCTTTCGGCGCGCGCGCTCAAGCG 293  
 Db 267 CAGCTATCCGAGCTGGCGCTCTTTCGGGGGCGCTTTCGGCGCGCGCGCTCAAGCG 326  
 Qy 294 CTGCAAGCAGCGCTGCGAGCCTTCCGCGAGTCCCGAGCCCGCGCGCGCGCTGCGCGGA 353  
 Db 327 CTGCAAGCAGCGCTGCGAGCCTTCCCGAGTCCCGAGCCCGCGCGCGCGCTGCGCGGA 386  
 Qy 354 CTTCCAGCGCGCGAGCCCTACAGTTTCTCGAGTTGCTTACTTCAAGCAAAATATCT 413  
 Db 387 CTTCCAGCGCGCGAGCCCTACAGTTTCTCGAGTTGCTTACTTCAAGCAAAATATCT 446  
 Qy 414 CCCCAAGCCATCGCGCTGCTCACACCTTCTTACTGAGCATCTCTGAGCAAAATGAT 473  
 Db 447 CCCCAAGCCATCGCGCTGCTCACACCTTCTTACTGAGCATCTCTGAGCAAAATGAT 506  
 Qy 474 GAAGAGCAATGCGCATATTAAGAGCCTGCTTGGTGGCGAGACTTACATTAAAGACCT 533  
 Db 507 GAAGAGCAATGCGCATATTAAGAGCCTGCTTGGTGGCGAGACTTACATTAAAGACCT 566  
 Qy 534 GGAACCAAGTCATATGAAGCCTGTTTCATCCGAGCAGTCCGCGCATACACGTCAGAA 593  
 Db 567 GGAACCAAGTCATATGAAGCCTGTTTCATCCGAGCAGTCCGCGCATACACGTCAGAA 626  
 Qy 594 CTGGAGAACTCCATCAGACATGGAGCTGGCGCTTCCGAGCTTCTTCAAGCCCTTTTA 653  
 Db 627 CTGGAGAACTCCATCAGACATGGAGCTGGCGCTTCCGAGCTTCTTCAAGCCCTTTTA 686  
 Qy 654 CGAGTGTCTCCAGCGCTGCGAGGCTTCCAGGAGATCAGGACTTCAAGGATTTTACCT 713  
 Db 687 CGAGTGTCTCCAGCGCTGCGAGGCTTCCAGGAGATCAGGACTTCAAGGATTTTACCT 746  
 Qy 714 TTCCATACAGATCATTATGTAGAGTTCTTGGAAATGCAAAATACAGTGTGAGAGAACCT 773  
 Db 747 TTCCATACAGATCATTATGTAGAGTTCTTGGAAATGCAAAATACAGTGTGAGAGAACCT 806  
 Qy 774 CACCCAGCTTATAGAGGCTATCCGTTGAGAAATTTGGCTACCATCTATCATTTACT 833  
 Db 807 CACCCAGCTTATAGAGGCTATCCGTTGAGAAATTTGGCTACCATCTATCATTTACT 866  
 Qy 834 GCAGTTTCCCTATTATTAAGTTGAACGACCTGGAAGATGAGAGCCCTGTGCGAGTACGTA 893  
 Db 867 GCAGTTTCCCTATTATTAAGTTGAACGACCTGGAAGATGAGAGCCCTGTGCGAGTACGTA 926  
 Qy 894 TCTGCTCTTTCATCAGAAATCAGAGGTCATCGCAGCAGAACCTGCTGTATTACCAGTACCA 953



Db 927 TCTGCTCTTTGATCAGATGACAGAGTCTATCAGCAGAACCTGCTGTATTACCACTACCA 986

Qy 954 CAGGACACTTGGGGCCCTCTCGATGACACTTCCAGCCAGACCTGAGCAGTTGAGTT 1013

Db 987 CAGGACACTTGGGGCCCTCTCGATGACACTTCCAGCCAGACCTGAGCAGTTGAGTT 1046

Qy 1014 CTTTAAATGTGACCACTCCAGAGAGGAGCTGTATGACTTTGCTTAAGGAAAATATATGGA 1073

Db 1047 CTTTAAATGTGACCACTCCAGAGAGGAGCTGTATGACTTTGCTTAAGGAAAATATATGGA 1106

Qy 1074 TGATGATGAGGAGAGAGTGTGTAATATGTGATGAGCTCTTGGAACTGGAGAGACCAAG 1133

Db 1107 TGATGATGAGGAGAGAGTGTGTAATATGTGATGAGCTCTTGGAACTGGAGAGACCAAG 1166

Qy 1134 C 1134

Db 1167 C 1167

BC049890 1663 bp mRNA linear ROD 10-JUN-2003

DEFINITION Mus musculus cartilage associated protein, mRNA (cDNA clone

ACCESSION MGC:54548 IMAGE:6306169), complete cds.

VERSION BC049890

KEYWORDS BC049890.1 GI:29437150

SOURCE MGC.

ORGANISM Mus musculus (house mouse)

REFERENCE 1 (bases 1 to 1663)

AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L.D., Shenmen, C.M., Schuler, G.D., Hopkins, R.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.D., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butcherfield, Y.S., Krzywinska, M.I., Skalska, U., Smalins, D.E., Schnerch, A., Schein, J.B., Jones, S.J., and Marra, M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 1663)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (31-MAR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Susan L. Sullivan, PhD.

cDNA Library Preparation: ResGen, Invitrogen Corp

DNA Library Arrayed by: The I.M.A.G.E. Consortium

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: [nisc\\_mgc@nhgri.nih.gov](mailto:nisc_mgc@nhgri.nih.gov)

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lari, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maekeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 99 Row: P Column: 3

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 9910169.

## FEATURES

## Location/Qualifiers

1..1663

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="MGC:54548 IMAGE:6306169"

/tissue\_type="Olfactory epithelium, neonatal mouse,

C57Bl/6<sup>+</sup>

/clone\_lib="NIH\_MGC\_129"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORTe"

1..1663

/gene="Crtp"

/notes="synonyms: CASP, 5730529N23Rik"

/db\_xref="LocusID:56693"

/db\_xref="MGI:1891221"

51..1253

/codon\_start=1

/product="cartilage associated protein"

/protein\_id="AAH49890.1"

/db\_xref="GI:29437151"

/db\_xref="LocusID:56693"

TRANSLATIONS "MGPRPTAALLVLLCGCAPTPGRGQYERYSFRFPRLMLP  
 SAYRALDOYSGHWAESVGLVLSLRHLRLRDLSEAFCHRCNSAATAPAPAGPASH  
 AEILRGLSVLRRAQCKRKCKQLPAPROSPRSVLADQREPQKFLQPAFYKFNKL  
 PKALAAHYLLKHPDEMKKNMEYKSLPGABDHIDLETKYESIESIFRVAFRAYNG  
 ENRATSDMELAPDLKAFYECIAACSGREIKDFKDFYLSIADHYVEVLECKIRC  
 EETLTPVIGYVPEKFAVMHYLQFAYYKLNDAAPCAVSYLLFQSDRVMQOOL  
 VYQYHRDKWGLSDEHFQRPRAVQFPNVTTLQKELYDPAQSHLMDDEGEVVEYVD  
 LLETSSA"

370 a 490 c 442 g 361 t

## BASE COUNT

## ORIGIN

Query Match 79.1%; Score 897.2; DB 10; Length 1663;

Best Local Similarity 87.0%; Pred. No. 9.4e-141;

Matches 983; Conservative 2; Mismatches 145; Indels 0; Gaps 0;

Qy 1 GGGCGCGCCCAATACGAACTACAGCTTCCGCGAGCTTCCACGGGACGAGCTGATGCGG 60

Db 117 GGGCGCGCGCAGTACGAGCGCTACAGCTTCCGCGAGCTTCCGCGGAGCGAGCTGCGG 176

Qy 61 CTCAGTTCGCGCTACCGGACCGCGCTGGACAGTACAGCGGCGAGACTGGGCCGAGAGC 120

Db 177 CTCAGTTCGCGCTACCGGACCGCGCTGGACAGTACAGCGGCGAGACTGGGCCGAGAGC 236

Qy 121 STKGCTTACCTGAGATACAGCTTCCGCGCTGCAACCGCTTCTGCGGACAGCGAGGCGCTTC 180

Db 237 GTGGCTACCTGAGATACAGCTTCCGCGCTGCAACCGCTTCTGCGGACAGCGAGGCGCTTC 296

Qy 181 TGGCACCGCACTGCGAGCGCGCGCGCGAGCGCGCGCGCGCGCGCTTCGCGAGCTAT 240

Db 297 TGGCACCGCACTGCGAGCGCGCGCGCGAGCGCGCGCGCGCGCGCTTCGCGAGCTAT 356

Qy 241 CCCGAGTTCGCGCTTCTGCGGCGCGCTGCTGCGCGCGCGCGCGCTTCGCGAGCTAT 300

Db 357 GCGCACTACGCTTCTGCGGCGCGCTGCTGCGCGCGCGCGCGCTTCGCGAGCTAT 416

Qy 301 CAGGCGCTTCGCGAGCTTCGCGCGAGTCCGCGAGTCCGCGAGTTCGCGCGAGCTTCGAG 360



```

Db 417 CAGGGCTGCGCGCTTCCGCCAGTCCGACCCAGCGGCTCAGTCTGGCGGACTTTTCAG 476
Qy 361 CGCGCGAGCCCTACAGATTCTGCGATTGCGTCTTCTCAGGCAAAATATCTCCCAAA 420
Db 477 CAGCGGAGCCCTACAGATTCTGCGATTGCGTCTTCTCAGGCAAAATATCTCCCAAG 536
Qy 421 GCATCGCGCTGCTGCTACAGCTTTCTCTGAAAGCTCTCTGATGAGCAAAATGATGAAGG 480
Db 537 GCATCGCTGCGCTGCTGCTACAGCTTTCTCTGAAAGCTCTCTGATGAGCAAAATGATGAAGG 596
Qy 481 AACATGCAATATTAAGAGCTTCTGCTGCGGAGGACTACATTAAGAGCTTGAAGG 540
Db 597 AACATGAGATTTAAGAGCTTCTGCTGCGGAGGACTACATTAAGAGCTTGAAGG 656
Qy 541 AAGTCATATGAAGCTTCTGCTGCGGAGGACTACATTAAGAGCTTGAAGG 600
Db 657 AAGTCGATGAGAGCTTCTGCTGCGGAGGACTACATTAAGAGCTTGAAGG 716
Qy 601 ACATCCATCAGACATGAGGCTTCTGCTGCGGAGGACTACATTAAGAGCTTGAAGG 660
Db 717 ACATCCATTTCCGACATGAGGCTTCTGCTGCGGAGGACTACATTAAGAGCTTGAAGG 776
Qy 661 CTGCGAGCTGCGGAGGCTTCCAGGAGATCAAGGACTTCAAGGACTTCTGAGGAGG 720
Db 777 CTGCGAGCTTCCGAGGAGGCTTCCAGGAGATCAAGGACTTCAAGGACTTCTGAGGAGG 836
Qy 721 GCAGATCAATATGAGGCTTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGG 780
Db 837 GCAGATCAATATGAGGCTTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGG 896
Qy 781 GTTATAGGAGGCTTCCGCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGG 840
Db 897 GTTATAGGAGGCTTCCGCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGG 956
Qy 841 GCTTATATAGGAGGCTTCCGCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGG 900
Db 957 GCTTATATAGGAGGCTTCCGCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGG 1016
Qy 901 TTTGATCAGATGAGGAGGCTTCCGCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGG 960
Db 1017 TTTGATCAGATGAGGAGGCTTCCGCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGG 1076
Qy 961 ACTTGGGGCTTCCGATGAGGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGG 1020
Db 1077 AAGTGGGGCTTCCGATGAGGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGG 1136
Qy 1021 GTGACCACTCCAGAGGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGG 1080
Db 1137 GTGACCACTCCAGAGGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGG 1196
Qy 1081 GAGGGAGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGG 1130
Db 1197 GAGGGAGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGGCTTCCGATGAGGAGG 1246

RESULT 10
MMAJ6469 1695 bp mRNA linear ROD 02-AUG-1999
LOCUS Mus musculus mRNA for cartilage-associated protein (CASP).
DEFINITION
ACCESSION AJ006469
VERSION AJ006469.1 GI:3687319
KEYWORDS cartilage-associated protein; CASP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
1 Morello, R., Tonachini, L., Monticone, M., Viggiano, L., Rocchi, M.,
Cancedda, R., and Castagnola, P.
TITLE cDNA cloning, characterization and chromosome mapping of Crtp
JOURNAL encoding the mouse cartilage associated protein
Matrix Biol. 18 (3), 319-324 (1999)
MEDLINE 99357019

```

```

PURNED 10429950
REFERENCE 2 (bases 1 to 1695)
AUTHORS Castagnola, P.
TITLES Direct Submission
JOURNAL Submitted (12-JUN-1998) Castagnola P., Biotechnologie in Oncologia,
Ist. Naz.le per la Ricerca sul Cancro, L.go R. Benzi 10 Genoa,
16132, ITALY
FEATURES
Location/Qualifiers
source 1..1685
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="embryo"
/dev_stage="14-18 d.p.c."
gene 1..1685
/genes="casp"
CDS 72..1274
/genes="casp"
/codon_start=1
/product="cartilage-associated protein (CASP)"
/protein_id="CAA07053.1"
/db_xref="GI:3687320"
/db_xref="SP:RMSL:088698"
/translation="MGPRSPAAALVLLVCVCAPTGRCQYERYFRNPPDELMPLF
AYRLHALDQYSGHEWAEVSYLEVLRLHLLRDSFAFCHRNCSAATAPAPAGPASH
AEALFGSVLRRAQCLKRCQGLPAPFQSPQSRVLAHQFQREPKYKFLQFAPFKNAL
PKATAAHTVLLKHPDDEMMKEMMYKSLPGAEDHDKLETKSYELFVFAVYKNG
ENWRTSISDMELALPDLKAFVCELAACGSRDKDFKDFVLSIADHVFVECKIRC
EETLTPVIGYVPEKFAVYVLOFAYVYKLNKLAAPCAVSVLLPDSRVMOQNL
VYCYHRDKWGLSDEHFQPRPEAVQFNVTTLQKELYDFAQHLMDDEGEVWEYVDD
LLTEESA"
sig_peptide 72..146
/mat_peptide 147..1271
BASE COUNT 362 a 503 c 449 g 371 t
ORIGIN
Query Match 79.08; Score 895.6; DB 10; Length 1695;
Best Local Similarity 86.94; Pred. NO. 1.7e-140; Indels 0; Gaps 0;
Matches 982; Conservative 2; Mismatches 146;
Qy 1 GGGCGCGCCCAATAAGAACGCTACAGCTTCCGAGCTTCCACGGAGCAGAGCTGATGCCG 60
Db 138 GGGCGCGGGCAGTACGAGCGCTACAGCTTCCGCAACTTCCCGCGGAGCAGCTGATGCCG 197
Qy 61 CTCGAGTCGCGCTACCGGACGCGCTGGACAAGTACAGCGGCGAGCAGCTGGCGCGAGAGC 120
Db 198 CTCGAGTCGCGCTACCGGACGCGCTGGACAAGTACAGCGGCGAGCAGCTGGCGCGAGAGC 257
Qy 121 STKGCTACCTGGAGATCAGCTTGGCGGCTGACCGCTTGTCTGGCGGACAGCGAGCCTTC 180
Db 258 GTGGGCTACCTGGAGGTCAGCTTGGCGGCTGACCGCTTGTCTGGCGGACAGCGAGCCTTC 317
Qy 181 TGCCACCGCAACTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGCCAGCTAT 240
Db 318 TGCCACCGCAACTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 377
Qy 241 CCCGAGTCGCGCTTTCGGGGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 300
Db 378 GCCGAACCTACGCTCTTCGGGACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 437
Qy 301 CAGGGCTTGCAGCGCTTCCGCCAGTCCCGAGCCAGCCGCGAGGTCGCTGGCGAGCTTCAG 360
Db 438 CAGGGCTTGCAGCGCTTCCGCCAGTCCCGAGCCAGCCGCGAGGTCGCTGGCGAGCTTCAG 497
Qy 361 CGCGCGAGCGCTTACAGTTCCTGAGTTCGCTTCTTCAAGGCAAAATATCTCCCAAA 420
Db 498 CAGCGGAGCGCTTACAGTTCCTGAGTTCGCTTCTTCAAGGCAAAATATCTCCCAAG 557
Qy 421 GCCATCGCGCTGCTCAGACCTTTCTTACTGAAGCATCTCTGATGAGCAAAATGATGAAGG 480

```

Db 558 GCCATCGCTCGGCTCACACCTATCTCTGAGCATCCAGATGACGAGATGATGAAGAGA 617  
 QY 481 AACATGGCATATATAAGAGCCTGCTGGTGCAGGAGTACATATAAGACCTGGAAC 540  
 Db 618 AACATGGAGTATATAAGAGCTTGCTGGAGCCGAGGACCAATATAAGACTTGGAAAC 677  
 QY 541 AAGTCATATGAAGGCTGTTCATCCGAGCAGTGGGCGATACACGCTGGAACCTGGAGA 600  
 Db 678 AAGTCGACGAGCCTGTTGCTCGCTGGTGGGCGTACACGCGGAGAACTGGAGA 737  
 QY 601 ACATCCATCACAGACATGAGCTGCCCTTCCGACTTCTTCAAGGCTTTTACGAGTGT 660  
 Db 738 AGCTCCATTTCCGACATGAGCTCGCTTCCGACTTCTTCAAGGCTTCTACGAGTGC 797  
 QY 661 CTCGAGCTCGAGGCTTCCAGGAGATCAGGACTTCAAGGATTTCTACCTTTCCATA 720  
 Db 798 CTGGCTGCTCGAGGCTCGCGGAGATCAAGACTTCAAGGACTTCTACCTGTCATA 857  
 QY 721 CGAGATCATATGTAGAAGTCTGGAATGCAAAATACAGTGTGAAGAGAACTCACCCCA 780  
 Db 858 GCAGATCACTATGTGAAGTCTGGAGTGAAGATTCGTTGTGAGGAGAGCCCTCACCCCA 917  
 QY 781 GTTATAGAGGCTATCCGCTTGAGAAATTTGCTGCTACCATGTATCATTTGCGATTT 840  
 Db 918 GTCATAGAGGCTATCCGCTGAGAAATTTGCGGACCATGTACCATTTTACAGTTT 977  
 QY 841 GCCTATTATTAAGTTGAACGACTGAAGAAATGACGCCCTGTGAGTCAGCTATCTGCTC 900  
 Db 978 GCGTATTACAAGTTGAATGATCTGAAGAAATGACGCCCTGTGCGCTCAGCTACTGCTC 1037  
 QY 901 TTTGATCAGAATGACAAGTCTATGACGAGACCTGGTGTATACCACTAGTACCAAGGAC 960  
 Db 1038 TTTGACCAAGTGTACAGGCTATGCAAGAACCTGTGTACTATCATAGTACCAAGGAC 1097  
 QY 961 ACTTGGGCTCTCGGATGAGCACTTCCAGCCGACCTGAAGCAGTTCAGTTCCTTTAAT 1020  
 Db 1098 AAGTGGGCTCTCGGATGAGCACTTCCAGCCGACCTGAGACCCGAGCAGTTCAGTTCCTTTAAT 1157  
 QY 1021 GTGACACACTCCAGAGGCTGATGACTTCTTCTAGGAAATATAAGTGGATGAT 1080  
 Db 1158 GTGACAGCTCCAGAGAACTGTACGACTTCCTCAGGAACACCTAATGGATGAGAT 1217  
 QY 1081 GAGGAGAGTGTGGAATATGTGATGACCTTGGAACTGGAGGAGAC 1130  
 Db 1218 GAGGAGAGTGTGAGTATGTGAGCACTTGTGAGACGGAAGATC 1267

## RESULT 11

AK097650

LOCUS

DEFINITION Homo sapiens cDNA FLJ40331 fis, clone TEST12031687, highly similar to Homo sapiens mRNA for cartilage-associated protein.

ACCESSION

AK097650

VERSION

AK097650.1

KEYWORDS

oligo capping; fis (full insert sequence).

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

1 Kanehori, K., Ishibashi, T., Chiba, Y., Fujimori, K., Hiraoka, S., Tanai, H., Watanabe, S., Ishida, S., Ono, Y., Horita, T., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and Isogai, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 1810)  
 Isogai, T. and Yamamoto, J.  
 Direct Submission

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

## JOURNAL

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); cDNA library  
 construction: Helix Research Institute (HRI) (supported by Japan  
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
 HRI, and Biotechnology Center, National Institute of Technology and  
 Evaluation; clone selection for full insert sequencing: HRI and  
 RAB; annotation: HRI and RAB.

## FEATURES

source

1. 1810  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="caxon:9606"  
 /clone="TEST12031687"  
 /tissue\_type="testis"  
 /clone\_lib="TEST12"  
 /note="cloning vector: pMB18SPL3"

BASE COUNT 375 a 550 c 449 g 436 t

## ORIGIN

Query Match 75.0%; Score 850; DB 9; Length 1810;  
 Best Local Similarity 88.2%; Pred. No. 7.7e-133;  
 Matches 1000; Conservative 2; Mismatches 2; Indels 130; Gaps 2;  
 QY 1 GGGCGCGCCCAATACGAACGCTCAGCTTCCGAGCTTCCACGGGACGAGCTGATGCCG 60  
 Db 94 GGGCGCGCCCAATACGAACGCTCAGCTTCCGAGCTTCCACGGGACGAGCTGATGCCG 153  
 QY 61 CTCGAGTCGGCTACCGCGACGCGCTGGACAAAGTACAGCGCGGACGACTGGGCGGAGAC 120  
 Db 154 CTCGAGTCGGCTACCGCGACGCGCTGGACAAAGTACAGCGCGGACGACTGGGCGGAGAC 213  
 QY 121 STKGCTACCTGGAGATCAGCTCGCGCTGCACCGCTTGCTGCGGACGACGAGGCTTC 180  
 Db 214 GTGGCTACCTGGAGATCAGCTCGCGCTGCACCGCTTGCTGCGGACGAGGCTTC 273  
 QY 181 TGCACCGCAACTGCGAGCGCGCGCGCGGACGCGCGCGCGCGCGCTTCGCGAGCTAT 240  
 Db 274 TGCACCGCAACTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCTTCGCGAGCTAT 333  
 QY 241 CCCGAGCTGCGCTTCTCGGGGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCTTCGCGAGCTAT 300  
 Db 334 CCCGAGCTGCGCTTCTCGGGGGCTGCTGCGCGCGCGCGCGCGCGCGCGCTTCGCGAGCTAT 393  
 QY 301 CAGGCGCTGCGAGCTTCCGCGAGTCCAGCGCCGACGCGCGCGAGGTGCTGGCGGACTTCCAG 360  
 Db 394 CAGGCGCTGCGAGCTTCCGCGAGTCCAGCGCCGACGCGCGCGAGGTGCTGGCGGACTTCCAG 453  
 QY 361 CCGCGCGAGCGCTTACAAAGTTCCTGCGAGTTCGCTTACTTCAAGGCAAAATATCTCCCAA 420  
 Db 454 CCGCGCGAGCGCTTACAAAGTTCCTGCGAGTTCGCTTACTTCAAGGCAAAATATCTCCCAA 513  
 QY 421 GCGATGCGCGCTGCTACACCTTTCTACTGAAGCATCTCGATGACGAATGATGAAGAG 480  
 Db 514 GCGATGCGCGCTGCTACACCTTTCTACTGAAGCATCTCGATGACGAATGATGAAGAG 573  
 QY 481 AACATGGCATATATAAGAGCCTGCTGGTGGCGAGGACTACATATAAGACCTTGGAAAC 540  
 Db 574 AACATGGCATATATAAGAGCCTGCTGGTGGCGAGGACTAC-TTAAAGACCTTGGAAAC 632  
 QY 541 AAGTCATATAAGACCTTGTTCATCCGAGCAGTGGGCGCATACAAACGGTGAGAACTGGAGA 600  
 Db 633 AAGTCATATAAGACCTTGTTCATCCGAGCAGTGGGCGCATACAAACGGTGAGAACTGGAGA 692  
 QY 601 ACATCCATCACAGACATGAGCTGGCGCTTCCGAGCTTCTTCAAGGCTTTTACGAGTGT 660  
 Db 693 ACATCCATCACAGACATGAGCTGGCGCTTCCGAGCTTCTTCAAGGCTTTTACGAGTGT 752  
 QY 661 CTCGAGCTGCGAGGCTTCCAGGAGATCAAGGACTTCAAGGATTTCTACTTTCCTTCCATA 720

Db 753 CTCGAGGCTGCGAGGGTTCAGGAGATCAGGACCTCAGAGATTTCTACCTTTCCATA 812  
 Qy 721 GCAGATCATATATAGTGAAGCTTTGGAATGCAAAATACAGTGTGAAGAGAACTTCACCCCA 780  
 Db 813 GCAG----- 816  
 Qy 781 GTTATAGGAGGCTATCCGGTTTGAGAAATTTGTGGCTACCATGTATCTACTTGCAGTTT 840  
 Db 817 ----- 816  
 Qy 841 GCCTATTATAGTTGAACGACTGAAGATGAGAGCCCTCTGAGTCAGTCATCTCTC 900  
 Db 817 -----TGAACGACCTGAGATGAGAGCCCTCTGAGTCAGTCATCTCTC 863  
 Qy 901 TTTGATCAGAAATGACAAGGTCTAGCAGCAGAACCTGTGTATTACCAAGTACCAAGGGAC 960  
 Db 864 TTTGATCAGAAATGACAAGGTCTAGCAGCAGAACCTGTGTATTACCAAGTACCAAGGGAC 923  
 Qy 961 ACTTGGGGCTCTCGATGAGCAGCTCCAGCCAGCCTGAGCAGTTCAGTTCTTTAAT 1020  
 Db 924 ACTTGGGGCTCTCGATGAGCAGCTCCAGCCAGCCTGAGCAGTTCAGTTCTTTAAT 983  
 Qy 1021 GTGACCACACTCCAGAGGAGCTGTATGACTTTTGTAAAGGAAATATATAGATGATGAT 1080  
 Db 984 GTGACCACACTCCAGAGGAGCTGTATGACTTTTGTAAAGGAAATATATAGATGATGAT 1043  
 Qy 1081 GAGGAGAGTGTGGAATATGTGATGAGCCTTTGGAAGTGGAGGAGCAGC 1134  
 Db 1044 GAGGAGAGTGTGGAATATGTGATGAGCCTTTGGAAGTGGAGGAGCAGC 1097

RESULT 12  
 BD148800  
 LOCUS  
 DEFINITION  
 Primer for synthesizing full-length cDNA and use thereof.  
 ACCESSION  
 BD148800  
 VERSION  
 BD148800.1 GI:27854558  
 KEYWORDS  
 JP 2002191363-A/3643  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Oca, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
 Primer for synthesizing full-length cDNA and use thereof  
 Patent: JP 2002191363-A 3643 09-JUL-2002;  
 HELIX RESEARCH INSTITUTE  
 OS Homo sapiens (human)  
 PN JP 2002191363-A/3643  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
 PI SAITO,  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC  
 C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
 10,  
 PC C12P21/02, C12Q1/68//C12P2/08, G06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key  
 Location/Qualifiers  
 FT source 1. 818  
 FT Location/Qualifiers  
 1. 818  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 159 a -271 c 236 g 148 t 4 others

Query Match

51.8%; Score 587.6; DB 6; Length 818;

Best Local Similarity 96.2%; Pred. No. 8.1e-89;  
 Matches 631; Conservative 2; Mismatches 18; Indels 5; Gaps 3;  
 Qy 1 GGGCGCCCAATAACGAAGCTTACAGCTTCCGAGCTTCCACGGAGAGAGTGTATGCCG 60  
 Db 109 GGGCGCCCAATAACGAAGCTTACAGCTTCCGAGCTTCCACGGAGAGAGTGTATGCCG 168  
 Qy 61 CTCAGTTCGGCTTACCGGACGCGCTGGACAAGTACAGCGCGAGCACTTGGGCGGAGAGC 120  
 Db 169 CTCAGTTCGGCTTACCGGACGCGCTGGACAAGTACAGCGCGAGCACTTGGGCGGAGAGC 228  
 Qy 121 STGGCTACCTGAGATCAGCTTCGGCTGACACCTGCTGCTGCGGAGAGAGAGCTTTC 180  
 Db 229 GTGGCTACCTGAGATCAGCTTCGGCTGACACCTGCTGCTGCGGAGAGAGAGCTTTC 288  
 Qy 181 TGGCACCGCAACTGACGCGCGCGCGAGCCGCGAGCGCGCGCGCTGCGCAGCTAT 240  
 Db 289 TGGCACCGCAACTGACGCGCGCGCGAGCCGCGAGCGCGCGCGCTGCGCAGCTAT 348  
 Qy 241 CCGAGCTGGCTTCCTCGGGGCTTCTCGGGGCTTCTCGCGCGCGCTGCTCAAGCGCTGCAAG 300  
 Db 349 CCGAGCTGGCTTCCTCGGGGCTTCTCGGGGCTTCTCGCGCGCGCTGCTCAAGCGCTGCAAG 408  
 Qy 301 CAGGCGCTGCGAGCTTCCGCGAGTCCAGCCGCGAGCGCGCGAGTGTGCGGAGCTTCCAG 360  
 Db 409 CAGGCGCTGCGAGCTTCCGCGAGTCCAGCCGCGAGCGCGCGAGTGTGCGGAGCTTCCAG 468  
 Qy 361 CGCGCGAGCGCTTACAGTTCCTGCGAGTTCGCTTCAAGGCAAAATATCTCCCCAAA 420  
 Db 469 CGCGCGAGCGCTTACAGTTCCTGCGAGTTCGCTTCAAGGCAAAATATCTCCCCAAA 528  
 Qy 421 GCCATCGCGCTGCTCACACCTTCTACAGTTCCTGAGTTCCTGAGTACGAAATGATGAGAGG 480  
 Db 529 GCCATCGCGCTGCTCACACCTTCTACAGTTCCTGAGTTCCTGAGTACGAAATGATGAGAGG 588  
 Qy 481 AACATGGCATATTATAAGAGCTTCCGCGAGTTCCTGCGGAGAGTACATTAAGAGCTTGAAGCC 540  
 Db 589 AACATGGCATATTATAAGAGCTTCCGCGAGTTCCTGCGGAGAGTACATTAAGAGCTTGAAGCC 647  
 Qy 541 AAGT--CATATGAAGCTTCTCATCGAGCTGCGGAGTACACGCTGAGAGTGA 598  
 Db 648 AAGTTCATATGAAGAGCTTTCATCGANCACTGCGGAGTACACGCTGAGAGTGA 707  
 Qy 599 GAACATTCATCACAGACATGAGCTGCGCC--TTCCCGAGCTTCTTCAAGAGCTTTT 652  
 Db 708 GAACATTCATCACAGACATGAGAGTGGCCCTTTCGAGTTCTTCAAGAGCTTT 763

## RESULT 13

GGCARPR  
 LOCUS  
 DEFINITION  
 G.gallus mRNA for cartilage associated protein.  
 ACCESSION  
 X97607  
 VERSION  
 X97607.1 GI:1296525  
 KEYWORDS  
 cartilage associated protein.  
 SOURCE  
 Gallus gallus (chicken)  
 ORGANISM  
 Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.

## REFERENCE

1  
 AUTHORS  
 Castagnola, P., Gennari, M., Morello, R., Tonachini, L., Marin, O.,  
 Gaggero, A. and Cancedda, R.  
 TITLE  
 Cartilage associated protein (CASP) is a novel developmentally  
 regulated chick embryo protein  
 J. Cell. Sci. 110 (Pt 12), 1351-1359 (1997)  
 J97360293  
 MEDLINE  
 PUBMED  
 9217321  
 REFERENCE  
 2 (bases 1 to 1665)  
 AUTHORS  
 Castagnola, P.  
 TITLE  
 Direct Submission  
 Submitted (29-APR-1996) P. Castagnola, Centro di Biotecnologie  
 Avanzate, Differenzamento Cellulare, L.go R. Benzi, 10 Genova,  
 JOURNAL

```

Db      739 TGTATAGCAGCGCTGTGAAGGCTCCCGAGAGATTAAAGATTTTAAAGACTTCATCTCTCT 799

QY      718 AFIGACGATCATTAATGATAGAAGTTCTGGAATGCAGAAAATACAGTGTTGA 764
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      799 ATTGCAGATCATTAATTAATTGAAGTCCTTGCATGCAAGTCAGTTCGA 845

RESULT 14
LOCUS   BC011701
DEFINITION Homo sapiens, nucleolar autoantigen (55kd) similar to rat synaptonemal complex protein, clone MGC:19699 IMAGE:3510378, mRNA, complete cds.
ACCESSION BC011701
VERSION    BC011701
KEYWORDS   MGC.
SOURCE     BC011701.1 GI:15079791
ORGANISM   Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1745)
Strausberg,R.
Direct Submission
Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M., Benjamin,B., Slakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R., Lim,M., Maduro,Q.L., Masello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J., Tongson,E.E., Touchman,J.W., Tsurgoun,C., Vogt,J.L., Walker,M.A., Zhang,L.-H. and Gresen,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 27 Row: d Column: 23
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.
Location/Qualifiers
1..1745
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocustID:10609"
/db_xref="taxon:9606"
/clone="MGC:19699 IMAGE:3510378"
/tissue_type="Brain, neuroblastoma"
/clone_lib="NIH MGC 19"
/lab_host="DH10B-R"
/note="Vector: pOTe7"
26..1339
/codon_start=1
/product="nucleolar autoantigen (55kd) similar to rat synaptonemal complex protein"
/protein_id="AAH1701.1"
/db_xref="GI:15079792"
/translaton="MARVAVGLMLLLGSAGAQVEKSPRGFPEDMLPAAAYGHAR EQYESWRREAFYEALRLRHLLRSEAFCHQSPPAKPQPDGGRAEWAC RYLLEFVRLEAACLRCKTLPAQVFPRQLLPQFSLEYLYHLALFGDAWAC"

FEATURES
source
```

Query Match  
Best Local Similarity 65.1%; Score 455.4; DB 9; Length 1745;  
Matches 723; Conservative 2; Mismatches 368; Indels 18; Gaps 3;

BASE COUNT 358 a 526 c 573 g 286 t

ORIGIN

5 GCGCCCAATACGAACTGCTCCGAGCTTCCCGAGCTTCCCGAGGAGAGCTGATCGCGCTCG 64  
65 AGTGGGCTACCGGACCGCTGGACAGTACAGCGGAGAGCTTGGGCGGAGAGCTG 124  
135 CCGCGGCTACCGGAGCGCTTGGAGCAGTACAGGAGAGAGCTGCGCGGAGAGCGCG 194  
125 GCTACCTGGAGATCAGCTGCGGCTGACCGCTTGTGGGAGACAGCGAGCGCTTCTGCC 184  
195 GCTACCTGGAGGCGGCTGCGGCTGACCGCTTGTGGGAGAGAGCGAGCGCTTCTGCC 254  
185 ACAGCACTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 235  
255 ACAGCACTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 314  
236 GCTA-----TCCCGAGCTGCGGCTTCCCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 289  
315 ACAGTGGGCTGCGGCTGCGGCTTCTCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 374  
290 AGCGCTGAGAGGCGCTGCGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCG 349  
375 GCGGCTGAGAGGCGGCTGCGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCG 434  
350 CGGACTTCCAGCGCGGCTGCGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCG 409  
435 GTGACTTCCAGCGCGGCTGCGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCG 494  
410 ATCTCCCAAGGAGCGGCTGCGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCG 469  
495 GCTGAGAGGCGGCTGCGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCT 554  
470 TGATGAGAGGAGATGCGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCT 526  
555 TGACCGCAAGTATCTCACTATCAGGAGGATGCTGAGCGGCTTCCCGAGCTTCCCGAGCTTCCCG 614  
527 AAGACCTGGAACCAAGTATGAAAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCG 586  
615 CGGACTGAGGCGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCG 674  
587 GTGAGAGTGGAGACATCCATCAGAGATGAGGCTGCGGCTTCCCGAGCTTCCCGAGCTTCCCGAG 646  
675 GCGGGGATTTCCCGAGAGCAGGAGATGAGGCGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCG 734  
647 CTTTTCAGAGTGTCTCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCT 706  
735 TCTTTCGCGGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAGCT 794  
707 TCTACCTTCCATAGCAGATCATATGAGAGTTCGAAATGCGAAATACAGTGTGAG 766  
795 TCTACCGGCGCATAGCAGATCTCTTTCGAGAGTCCCTGCGAGTGCAGGTGAGCTGTGAGG 854  
767 AGAACCTCACCAGTATATAGGAGGCTATCCGCTTCCCGAGCTTCCCGAGCTTCCCGAGCTTCCCGAG 826  
855 CCAATTTGACCGCAATGCTGGTGGCTTCTCGTGGAGCAAGTTCGTCGCGCAGCTGATGAC 914  
827 ATTACTTGCATTTGCTTATATAGTGAACCTGAGGATGAGGCTGAGGCTGAGGCTGAGGCTGAGG 886  
915 ACTACCTGAGTTTTCCTTATATAGTGAATGAGTGGCGAGGCTGAGGCTGAGGCTGAGGCTGAGG 974  
887 TCAGTATCTGCTTTTGTATGAGATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 946

Db 975 CCAGTACATGCTCTTCGACCCCGAGGACAGGCTCATGCGAGCAACCTGGTATTACC 1034  
Qy 947 AGTACCACAGGACACTTGGGGCTCTCGGATGAGCAGCTTCCAGCCAGACCTGAGCAG 1006  
Db 1035 GTTTCACCGGGCTCGCTGGGCTTGAAGAGAGGAGCTTCCAGCCCGGGAGGAGGCA 1094  
Qy 1007 TTAGTCTTTTATGATGACCACTCCAGAGAGCTGTATGACTTTGCTAGGAAATA 1066  
Db 1095 TGCTTACCAACACAGAGCGGAGCTCGGAGCTGCTGGAGTTCACCCACATGTACC 1154  
Qy 1067 TAATGGATGATGATGAGGAGAGTGTGTGGA 1097  
Db 1155 TGCAGTCAATGATGATGAGTGGAGCTGGAGGA 1185

RESULT 15  
BC001047 1889 bp mRNA linear PRI 12-JUL-2001  
LOCUS Homo sapiens, nucleolar autoantigen (55kD) similar to rat  
DEFINITION synaptonemal complex protein, clone MGC:1322 IMAGE:3510406, mRNA,  
complete cds.  
ACCESSION BC001047 GI:12654442  
VERSION BC001047.1  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1889)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (11-DEC-2000) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
Contact: [amandansystemsbiology.org](mailto:amandansystemsbiology.org)  
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia  
Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IPAL Plate: 7 Row: f Column: 22  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 5454037.

## FEATURES

Location/Qualifiers  
1..1889  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="LocusID:10609"  
/db\_xref="taxon:9606"  
/clone="MGC:1322 IMAGE:3510406"  
/tissue\_type="Brain, neuroblastoma"  
/clone\_id="NIH MGC\_19"  
/lab\_host="DH10B-R"  
/note="vector: pOTB"  
47..1360  
/codon\_start=1  
/product="nucleolar autoantigen (55kD) similar to rat  
synaptonemal complex protein"  
/protein\_id="AAH01047.1"  
/db\_xref="GI:12654443"  
/translation="MARVAGLLMLLGGAGQYKYSRPFPPDMLPLAAYGHAL  
FOYEGSWBSAVLLENLELLEDSFCHANCSPAPAAKPDGGRADWACE  
LRLGRVLEAACLRCKRTLPAPVPPRLLRDFQSLPYQVHLVAFKALEK  
AVAAAYFLQRNFHETATKLYLTYQGLMDVADESITDLEAQPYEAVFLAVLYNSG

## CDS

DFRSSTEDMERALSEVLAVFARCLAGCEGAHQVDKDFYPAIADLFAELSLQCKVDC  
 ANLTPNVGVYFVDKFAVTHYHLYQFAYYKLANVROARSAASVYMLFPKDSVMQONLV  
 YVFFHARGLBEEDFQPEEAMLYHNTAEIRLELLETHMYLQSDDEMELEETEPEPL  
 EPDALSDAEFEGEGYBEGMVDWQSPDARKDEAEAPPELA"  
 BASE COUNT 399 a 570 c 611 g 309 t  
 ORIGIN

Query Match 40.2%; Score 455.4; DB 9; Length 1889;  
 Best Local Similarity 55.1%; Pred. No. 1.1e-66;  
 Matches 723; Conservative 2; Mismatches 368; Indels 18; Gaps 3;

Qy 5 GCGCCCAATACGAACGCTACAGCTTCCGAGCTTCCACGGGACGAGCTGATCGGCTCG 64  
 Db 96 GGGCGAGTACGAGAAGTACAGCTTCCGGGCTTCCGCCCGGAGGACCTGATCGGCTGG 155

Qy 65 AGTCGGCTACCGGACGCTGACAGTACAGGGGCGGAGCACTGGSCCGGAGGCTGK 124  
 Db 156 CCGCGGCTACGGGCGAGCTTGGAGAGTACAGGAGAGAGCTGGCGGAGAGCGCGC 215

Qy 125 GCTACCTGGAGATCAGCTCGGCTCCACCGCTTGTGCGGACAGCGAGGCTTCTGCC 184  
 Db 216 GCTACCTGGAGGCGGCTGCGGCTGACCGGCTTCCGCGGACAGCGAGGCTTCTGCC 275

Qy 185 ACGCCAATCGAGGCTGCGGCTTCCGCGGCGGAGCGGCGGCGGCGGCGGCGGCGCA 235  
 Db 276 ACGCCAATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 335

Qy 236 GCTA-----TCGCGAGCTGCGGCTTTCGGGGGCTGTCGGCGGCGGCGGCGGCGGCGG 289  
 Db 336 ACAGTGGGCTGCGAGCTGCGGCTTTCGGGGGCTGTCGGCGGCGGCGGCGGCGGCGG 395

Qy 290 AGCGCTGCAAGAGGCGGCTGCGGCTTCCGCGGCGGAGCGGCGGCGGCGGCGGCGGCGG 349  
 Db 396 GCGCTGCAAGGCGGCTGCGGCTTCCAGGTCGCTTCCGCGGCGGCGGCGGCGGCGGCGG 455

Qy 350 CGGACTTCCAGCGGCGGCGGCGGCTTCCAGTTCGCTGAGTTCGCTTCTTCAAGGCAATA 409  
 Db 456 GTGACTTCCAGAGCGGCTGCGGCTTCCAGTTCGCTGAGTTCGCTTCTTCAAGGCTACC 515

Qy 410 ATCTCCCGCAAGCGCTGCGGCTGCTCACACCTTCTTCAAGGAGCTGCTGATGACGAAA 469  
 Db 516 GGCTGGAGAGGCGGCTGCGGCGGCTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 575

Qy 470 TGATGAGAGGAGACATGGCAATATTA---GAGCGCTGCTGTCGGCGGCGGCGGCGGCGG 526  
 Db 576 TGACCGCAAGTATCAACTATCAAGGAGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGG 635

Qy 527 AAGACCTGGAACCAAGTCAATATGAAGGCTGTTTATCCGAGCAGTGGGGCGGCGGCGGCGG 586  
 Db 636 CGGACCTAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 695

Qy 587 GTGAGACTGGAGAACATCCATCACAGATGGAGTGGCGGCTTCCGCGGCTTCTTCAAG 646  
 Db 596 GCGGGGATTCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 755

Qy 647 CTTTTACGAGTGTCTCGAGCTGCGAGGCTTCCAGGAGGAGTCAAGGACTTCAAGGATT 706  
 Db 756 TCTTTGCGGCTGCTGCGGCGGCTGTAAGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 815

Qy 707 TCTACCTTCCATAGGAGTCAATATGTAAGTTCGGAATGCAAAATACAGTGTGAG 766  
 Db 816 TCTACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 875

Qy 767 AGAACCTTCAAGGAGTATAGGAGGCTATCCGGTGGAGAAATTTGTGGCTACCATGTATC 826  
 Db 876 CCAATTTGAGCCCGCAATGTGGTGGCTTCTTGTGGCAAGTTCTGTCGCGGCGGCGGCGG 935

Qy 827 ATTACTTGCAGTTTGCCTATTATAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 886  
 Db 936 ACTACTTGCAGTTTGCCTATTATAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 995

Qy 887 TCAGCTATCTGCTTTTGATCAGATGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 946

Search completed: January 30, 2004, 14:49:21  
 Job time : 4134.45 secs

Db 996 CCAGCTACATGCTCTTCGACCCCAAGGACAGCGCTCATGCGAGCAACCTGGTGTATTACC 1055  
 Qy 947 AGTACCAAGGAGGACACTTGGGGGCTCTTCGGATGAGCACTTCCAGCGGCGGAGCGAG 1006  
 Db 1056 GGTTCACCGGGCTCGCTGGGGCTTGAAGAGGAGGACTTCCAGCGGCGGAGGAGGCGCA 1115  
 Qy 1007 TTCAGTTCTTTAATGTGACCACTCCAGAGGAGGCTGTATGACTTTGCTTAAGGAAATA 1066  
 Db 1116 TGCTCTACCAACAGACCGCGAGCTGGGAGGCTTGGAGATTACCCACATGTATACC 1175  
 Qy 1067 TAATGATGATGATGAGGAGGAAAGTTGTGA 1097  
 Db 1176 TGCACTCAGATGATGAGATGAGGCTGGAGGA 1206